



# Etude de la stabilité de la méthylation ADN chez *Arabidopsis thaliana* et impact sur la transcription

Mélanie Rigal

## ► To cite this version:

Mélanie Rigal. Etude de la stabilité de la méthylation ADN chez *Arabidopsis thaliana* et impact sur la transcription. Sciences agricoles. Université Blaise Pascal - Clermont-Ferrand II, 2014. Français. NNT : 2014CLF22496 . tel-01378908

**HAL Id: tel-01378908**

**<https://theses.hal.science/tel-01378908>**

Submitted on 11 Oct 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE CLERMONT-FERRAND II – UNIVERSITE BLAISE PASCAL

N° D.U. 2496

ECOLE DOCTORALE DES SCIENCES DE LA VIE, SANTE, AGRONOMIE,  
ENVIRONNEMENT

N° d'Ordre 641

# THESE

Présentée pour obtenir le grade de

**DOCTEUR D'UNIVERSITE**

(spécialité Génétique et physiologie moléculaires)

## **Etude de la stabilité de la méthylation ADN chez *Arabidopsis thaliana* et impact sur la transcription**

Présentée par

**Mélanie RIGAL**

Soutenue le 10 Octobre 2014

**Directeur de thèse : Olivier MATHIEU**

### **MEMBRES DU JURY :**

**Examineur :** Nicolas BOUCHE (Chargé de recherche, INRA, Versailles)

**Rapporteur :** Jean-Marc DERAGON (Professeur, Université de Perpignan Via Domitia, Perpignan)

**Examineur :** Maria GALLEG0 (Professeur, Université Blaise Pascal, Clermont-Ferrand)

**Rapporteur :** Thierry LAGRANGE (Directeur de recherche CNRS, Perpignan)

**Examineur :** Marie MIROUZE (Chercheur IRD, Montpellier)



## Résumé

La maintenance de la méthylation ADN sur les sites CG joue un rôle crucial dans le silencing des éléments transposables (TE) et l'expression correcte des gènes. Chez *Arabidopsis thaliana*, on observe, dans le mutant *met1-3* déficient en méthylation CG, une apparition ectopique de méthylation CHG sur de nombreux gènes, ainsi qu'une relocalisation de la diméthylation de la lysine 9 de l'histone H3 (H3K9me2) de l'hétérochromatine vers l'euchromatine. Nous avons démontré que ceci est lié à un défaut de transcription au niveau du grand intron du gène codant la déméthylase H3K9me2 IBM1. Nous avons également constaté que dans les épihybrides F1 issus du croisement de plantes *met1-3* avec une plante sauvage la méthylation CHG présente dans l'intron de l'allèle *IBM1* hérité du parent mutant était perdue. Afin de définir si la perte de méthylation affecte également d'autres loci génomiques, et plus globalement à l'échelle du génome entier l'impact de la rencontre de deux épigénomes différents, nous avons analysé les profils de méthylation, de siRNAs et de transcription des épihybrides F1. Nos données révèlent que l'union de deux méthylomes distincts au sein d'un même génome provoque une restructuration considérable des profils épigénétiques et transcriptionnels. La méthylation CHG apparaissant sur de nombreux gènes dans *met1* tend à persister, créant ainsi de nouveaux épiallèles pouvant être hérités. Du côté des TE, nombre d'entre eux sont déméthylés et réactivés, tandis que d'autres sont immédiatement reméthylés et resilencés. Ainsi, ces résultats contribuent à la compréhension de la stabilité de la méthylation ADN et de son rôle dans le contrôle différentiel des gènes et des TE.

## Abstract

### Study of DNA methylation stability in *Arabidopsis thaliana* and impact on transcription

Maintenance of DNA methylation at CG sites is crucial for silencing of transposable element (TE) and proper expression of genes. In *Arabidopsis thaliana*, *met1-3* mutant, deficient in CG methylation, shows ectopic appearance of CHG methylation at numerous genes, as well as relocation of H3K9me2, from heterochromatin towards euchromatin. We have shown that this is due to a defect of the transcription of the large intron of the gene encoding the IBM1 H3K9me2 demethylase. We also found that, in the F1 epihybrids from the cross between *met1* and WT plants, CHG methylation at the intron of the *met1*-derived *IBM1* allele is lost. In order to define whether the loss of methylation at *IBM1* also affects other genomic loci, and the impact of the union of two different epigenomes on methylation and transcription genome-wide, we analyzed the methylation, siRNA and transcription patterns of F1 epihybrids. Our data reveal that the union of two distinct methylomes within the same genome triggers considerable restructuring of epigenetic and transcriptional patterns. CHG methylation appearing in the mutant parent tends to persist in F1, creating new epialleles that can be inherited. On the TE side, lots of them are demethylated and reactivated while others are immediately remethylated and resilenced. Thus, our results provide new insights to the understanding of DNA methylation stability and its role in the differential control of genes and TEs.

**Mots-clés :** *Arabidopsis*, épigénétique, méthylation ADN, MET1, hybrides, IBM1, histone, H3K9me2, transcription, siRNA.

**Keywords :** *Arabidopsis*, epigenetic, DNA methylation, MET1, hybrids, IBM1, histone, H3K9me2, transcription, siRNA.





## Remerciements :

Je tiens avant tout à remercier **Olivier Mathieu** de m'avoir permise de réaliser ces quatre années dans son laboratoire. Je n'aurai pu imaginer meilleur directeur de thèse, à la fois patient et exigeant, obstiné et minutieux, et toujours accessible, formant le petit Padawan aux techniques, à l'interprétation des données obtenues, m'incitant à me remettre un peu en question, mais pas TOUT le temps.

Je remercierai aussi mes compagnes et compagnons de labeur,

**Zoltan**, auprès de qui j'ai commencé cette thèse, dont j'ai pu apprécié le tempérament hongrois, les paroles réconfortantes visant à calmer mes angoisses...

**Yoko**, my Japanese sister, with whom I had a lot of great time, in the lab and outside, keeping me company late in the evening and the week-ends...

Ma pretty woman, **Leticia**, plus angoissée encore que moi, mais toujours à l'écoute et avec un cœur immense dans une taille de guêpe.

**Angeles**, dont l'arrivée a valu une belle escapade en girl trio à Lyon et dont j'ai pu apprécié le tempérament zen et de feu en même temps, locura español...

Les garçons aussi, **Steve**, **Anthony**, en mode iPad à fond, l'arrivée des « Montpelliérains », **Maxime** puis **Pierre**, et **Sébastien**, qui ont chacun fait ou font leur petit bonhomme de chemin dans l'équipe...

Un énorme remerciement d'ailleurs à **Pierre** pour le choix d'un concert merveilleux, une affiche de thèse Indiana Jonesque, d'une gentillesse et intelligence scientifique rare...

Un Grand Merci tout particulier à **Romain** aussi, dont la venue a bien facilité nos analyses bioinformatiques, mais a bien égayé également le quotidien.

Merci aux 2 Pélistier, **Mr Péloché**, qui m'a donné moult conseils techniques en tentant de garder patience quand j'ouvrais des yeux ronds et lui demandais de répéter pour la 4<sup>ème</sup> fois...

Un Immense merci aussi à **Anne-Sophie**, gestionnaire mais pas seulement, toujours à l'écoute, hyper efficace et adorable, à qui je ne peux que souhaiter d'atteindre une vie professionnelle stable et épanouissante.

Je remercie bien sûr aussi **tous les autres membres du laboratoire** que j'ai pu croiser durant ces quatre années de thèse, les étudiants comme les chercheurs des trois équipes du couloir.

Je remercierai aussi très sincèrement **Detlef Weigel** et **Claude Becker** pour notre collaboration, m'avoir accueilli dans le laboratoire, formé au séquençage ARN puis avoir contribué à l'analyse des données et à la rédaction du manuscrit.



Je dédierai ce manuscrit à **Ma Famille**, mes parents, mes frère et sœur et tous les autres, ayant subi mes absences chroniques, mon obsession pour le labo, mes angoisses et ma presque volatilisation durant ces quatre années... et Un Immense

Merci à mes petits choux, **neveu et nièce**, ayant décidé d'arriver pendant cette période et m'ayant ainsi offert des instants de bonheur, de rigolade et de maternage salvateurs.

Je remercierai aussi **mes amies** qui m'ont, je crois, pardonné mon absence et mon obsession, durant ces années, et qui m'ont suivi jusqu'à aujourd'hui...

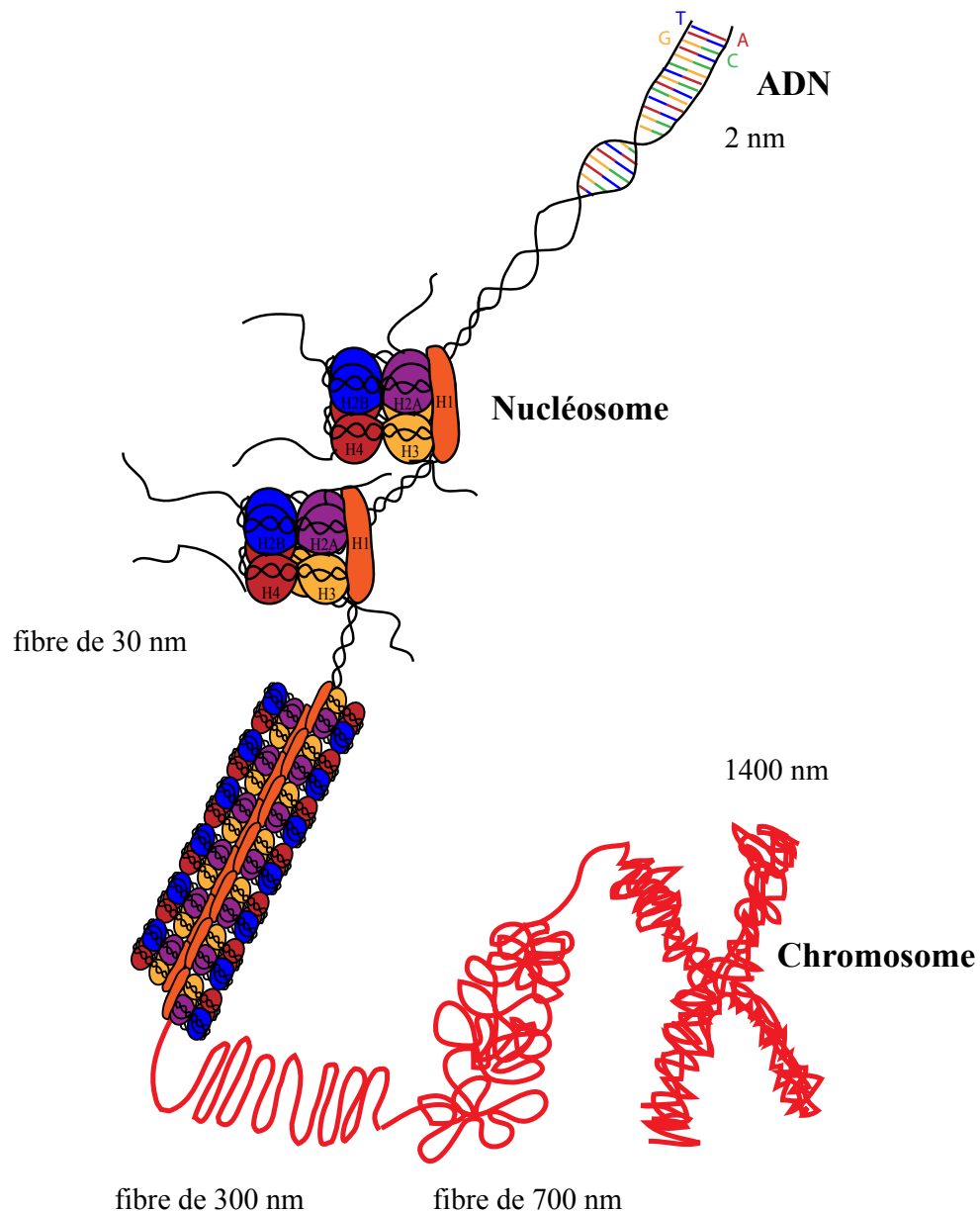
En tout dernier point, je souhaite remercier les **membres du jury** d'avoir accepté notre invitation, lu le manuscrit dense et fastidieux durant les beaux jours d'été, et offert à ma soutenance des remarques qui nous ont particulièrement été utiles notamment pour la suite des analyses de séquençage et de la rédaction de l'article.

Merci à Vous ...



# **TABLE DES MATIERES**

<b>I. INTRODUCTION .....</b>	<b>2</b>
<b>I.A. Voies épigénétiques chez <i>Arabidopsis thaliana</i> .....</b>	<b>2</b>
I.A.1. L'épigénétique.....	2
I.A.2. Composition du génome d' <i>Arabidopsis thaliana</i> .....	4
I.A.3. La transcription chez <i>Arabidopsis thaliana</i> .....	6
I.A.4. Répression des éléments transposables .....	8
I.A.5. Régulation épigénétique des gènes .....	24
I.A.4.a. La méthylation ADN.....	24
I.A.4.b. La déméthylation ADN.....	30
I.A.4.c. Les marques histones et variants d'histones.....	36
I.A.4.a.1. L'acétylation et la déacétylation des histones .....	36
I.A.4.a.2. L'ubiquitination et déubiquitination des histones.....	38
I.A.4.a.3. La méthylation/déméthylation des histones.....	40
I.A.4.a.4. Les variants d'histones.....	48
I.A.4.b. La voie RNA-directed DNA methylation (RdDM) .....	50
I.A.4.c. Rôle de la méthylation CG sur les gènes.....	52
<b>I.B. Stabilité de la méthylation ADN au cours des générations.....</b>	<b>56</b>
I.B.1. Les épiallèles naturels .....	56
I.B.2. Variabilité épigénétiques naturelle.....	60
I.B.2.a. Profils épigénétiques des populations naturelles.....	60
I.B.2.b Impact de la formation d'hybrides sur les profils épigénétiques.....	64
I.B.3. Epiallèles induits dans des mutants hypométhylés .....	68
I.B.4. Rétrocroisement de mutants hypométhylés .....	70
I.B.5. Analyse de lignées épigénétiques recombinantes .....	72
<b>I.C. Objectifs de la thèse .....</b>	<b>76</b>
 <b>II. CHAPITRE 1 : ETUDE DE LA REGULATION DU GENE <i>IBM1</i></b>	
<b>DANS LE MUTANT <i>MET1</i> .....</b>	<b>78</b>
<b>II.A. Introduction .....</b>	<b>80</b>
<b>II.B. Article « Rigal et al., 2012 » .....</b>	<b>82</b>
 <b>III. CHAPITRE 2 : CONSEQUENCES IMMEDIATES DE LA</b>	
<b>RENCONTRE DE PLANTES AUX GENOMES IDENTIQUES MAIS</b>	
<b>AUX EPIGENOMES DISTINCTS – DYNAMIQUE DE LA</b>	
<b>METHYLATION ADN ET DE LA TRANSCRIPTION DANS LES</b>	
<b>EPIYBRIDES F1.....</b>	<b>132</b>
<b>III.A. Introduction.....</b>	<b>134</b>
<b>III.B. Article « Rigal et al., 2015 ».....</b>	<b>138</b>
 <b>IV. DISCUSSION GENERALE .....</b>	<b>234</b>
 <b>RÉFÉRENCES BIBLIOGRAPHIQUES.....</b>	<b>248</b>



**Figure I-1 : Représentation schématique de la structure de la chromatine.**

La molécule d'ADN d'un diamètre de 2 nm est enroulée tout d'abord sur les octamères d'histones comportant deux dimères H2A-H2B et deux dimères H3-H4. Ceci forme le nucléosome. De plus, cette structure est stabilisée par la liaison de l'histone H1 à l'ADN internucléosomal, permettant le repliement du nucléofilament en fibre de 30 nm, elle-même compactée en boucles de 300 nm d'envergure. Ces boucles sont organisées en super-hélices formant ainsi la structure en chromosome de 1400 nm.

## I. INTRODUCTION

### I.A. Voies épigénétiques chez *Arabidopsis thaliana*

#### I.A.1. L'épigénétique

La découverte de l'ADN comme support de l'information génétique par Watson et Crick il y a plus de 60 ans a représenté une avancée majeure dans la compréhension des mécanismes impliqués dans le développement et le fonctionnement des organismes vivants. Cependant, un mystère demeurait sous-jacent : comment toutes les cellules d'un individu peuvent-elles porter la même information génétique mais avoir des caractéristiques particulières, au cours du développement mais aussi selon les organes, les tissus ?

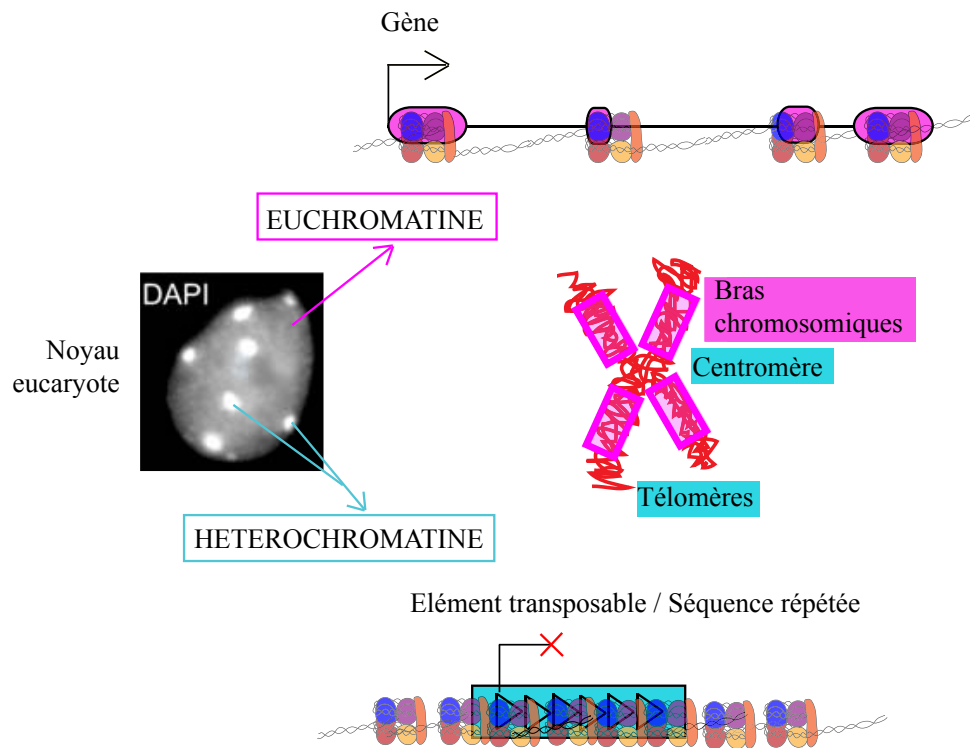
L'épigénétique a permis de lever le voile au moins partiellement sur ce paradoxe. La notion d'épigénétique est définie, dès les années 1950, comme l'étude des mécanismes moléculaires participant à la régulation de l'expression des gènes au cours du développement et de la différenciation cellulaire. Les modifications épigénétiques n'affectent pas la séquence d'ADN elle-même mais peuvent être transmises d'une génération à l'autre et être réversibles. De plus, elles peuvent être modulées par des facteurs environnementaux. Comme détaillé plus en avant dans les paragraphes suivants, elles jouent également un rôle crucial dans la régulation des éléments transposables, composants majeurs des génomes eucaryotes dont la mobilisation, pouvant engendrer des réarrangements chromosomiques et des mutations, menace l'intégrité du génome.

#### **Que ciblent ces modifications épigénétiques ?**

L'ADN est fortement compacté dans le noyau grâce à de nombreuses protéines, ce qui forme la chromatine, structure cible des modifications épigénétiques (**Figure I-1**).

Le premier niveau de compaction de l'ADN consiste en l'enroulement de la molécule autour d'octamères d'histones (deux dimères H3-H4 et deux dimères H2A-H2B), formant les nucléosomes (Luger et al. 1997). Chaque nucléosome contient environ 145 pb d'ADN et est séparé du nucléosome voisin par un fragment d'ADN de liaison de 10 à 80 bp. L'assemblage de ces histones dans la chromatine a lieu le plus souvent lors de la réplication, et implique des chaperones d'histones, telles que CAF-1.





**Figure I-2 : Deux types de chromatine sont retrouvés dans les noyaux eucaryotes.**

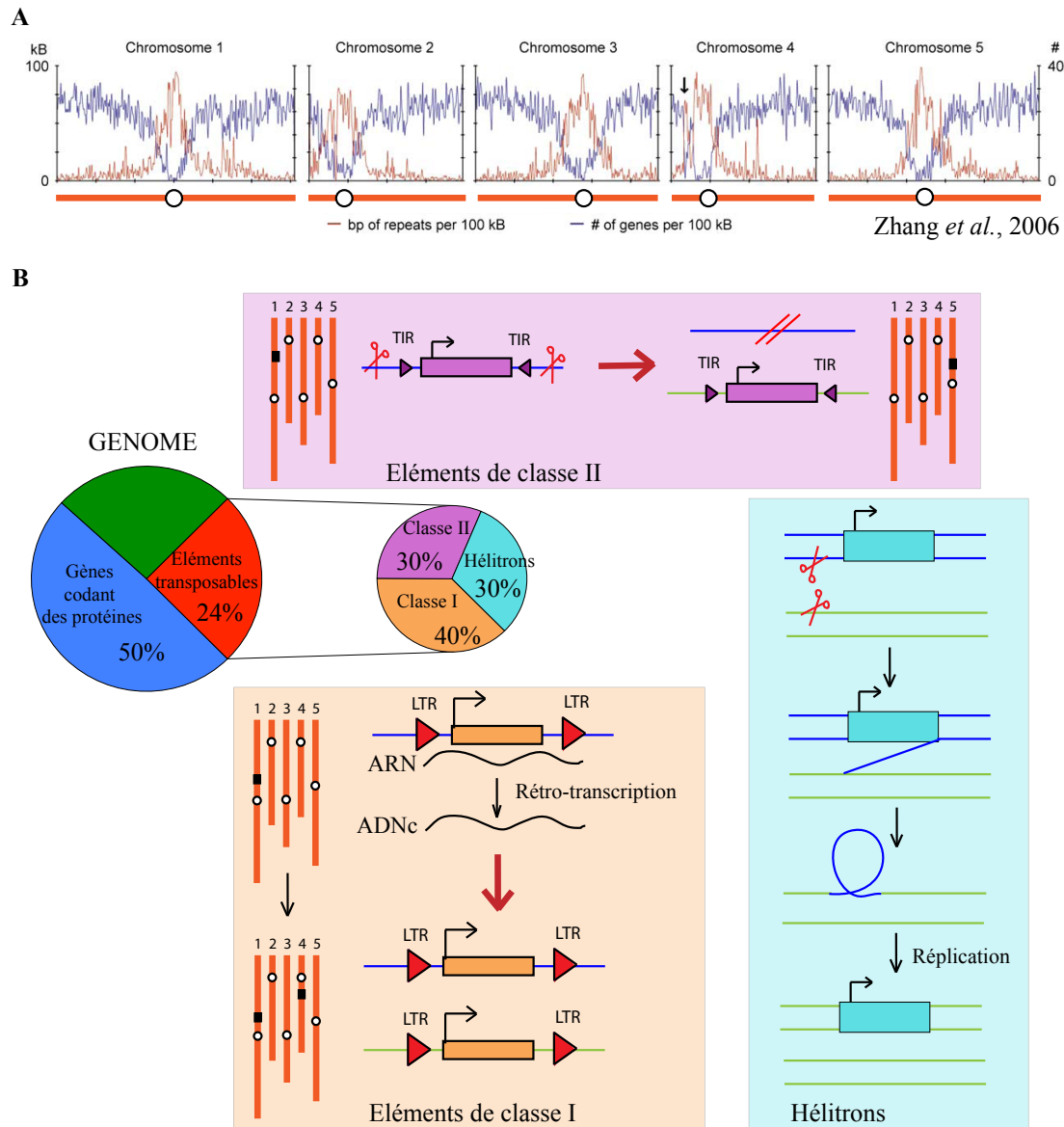
L'euchromatine, peu condensée est riche en gènes et localisée principalement sur les bras chromosomiques chez *A. thaliana*. L'hétérochromatine, quant à elle, est très condensée, riche en histones H1 et enrichie en éléments transposables et séquences répétées, elle est retrouvée dans les zones péri-centromériques et télomériques.

Secondairement, l'histone H1 s'associe à cet ADN de liaison et permet ainsi d'augmenter le degré de compaction et former une fibre de 30 nm. La fibre chromatinienne se condense à la mitose sous forme de chromosomes. Les analyses cytologiques mettent en évidence deux types de chromatine : l'euchromatine, peu condensée, riche en gènes, et l'hétérochromatine, très condensée grâce à la présence de l'histone H1, incluant chez la plante modèle *Arabidopsis thaliana* les régions télomériques, et péricentromériques (**Figure I-2**).

### **I.A.2. Composition du génome d'*Arabidopsis thaliana***

La plante modèle *Arabidopsis thaliana* possède un génome de 125 Mb, se répartissant sur cinq chromosomes ; il s'agit d'un des plus petits génomes du règne végétal d'après les données actuelles (2000). Les gènes codant des protéines occupent 50% du génome. Au nombre de 27 411, ils ont une taille moyenne de 2 kb, avec des exons de 296 pb et des introns de 165 pb en moyenne. Les éléments transposables représentent, eux, 24% du génome, une proportion faible en comparaison à d'autres plantes comme le Maïs par exemple (85% du génome) (**Figure I-3A&B**). Cette portion du génome inclut à 40% des éléments de classe I, qui transposent par un mécanisme de « copier/coller » via un intermédiaire ARN. Ces éléments sont majoritairement retrouvés au niveau des centromères.

Les 60% restants sont occupés en part égale par des transposons de classe II, se déplaçant, eux, par un mode de « couper/coller », et des Hélitrons, dont le mode de transposition par cercle roulant est encore mal connu. Ces transposons ont une distribution plus large et variable que les éléments de classe I, même si fortement regroupés dans les zones péricentromériques et les « knobs » hétérochromatiques. Les Miniature Inverted repeat Transposable Elements (MITEs), hobo, Activator, Tam3 (hAT) et Hélitrons sont aussi largement retrouvés sur les bras chromosomiques, riches en gènes (2000).



**Figure I-3 : Distribution du génome d'*Arabidopsis thaliana* en gènes et éléments transposables.**

**A-** Chez *A. thaliana*, sur les cinq chromosomes, les séquences répétées sont préférentiellement localisées dans les zones péricentromériques tandis que les gènes sont plus présents sur les bras chromosomiques.

**B-** Les gènes occupent la moitié du génome d'*A. thaliana* et les éléments transposables près d'un quart. Les éléments de classe I transposent par un mécanisme de « copier/coller » via un intermédiaire ARN, ainsi leur mobilisation provoque une augmentation de leur nombre de copies. Les éléments de classe II se déplacent par un mode de « copier/coller », changeant ainsi de localisation génomique. Les Hélitrons, eux, sont mobilisés via un mode de transposition peu connu à ce jour, qui mettrait en jeu une structure en cercle roulant.

### I.A.3. La transcription chez *Arabidopsis thaliana*

Chez *Arabidopsis*, cinq polymérases ARN ont été identifiées et caractérisées (**Figure I-4A**). La polymérase I (Pol I) est la plus spécialisée, elle transcrit les gènes codant les ARN ribosomiaux (ARNr) 18S, 5.8S et 28S. La polymérase II (Pol II) transcrit un très grand nombre de gènes codant des protéines mais aussi des ARN non-codants (ARNnc) et des petits ARN modulant ou maturant les ARN messagers (ARNm) ou les ARNr. Enfin, la polymérase III (Pol III) transcrit diverses classes de petits ARN tels que les ARN de transfert (ARNt), l'ARNr 5S, de petits ARN régulateurs ainsi que les Short Interspersed Nuclear Elements (SINE).

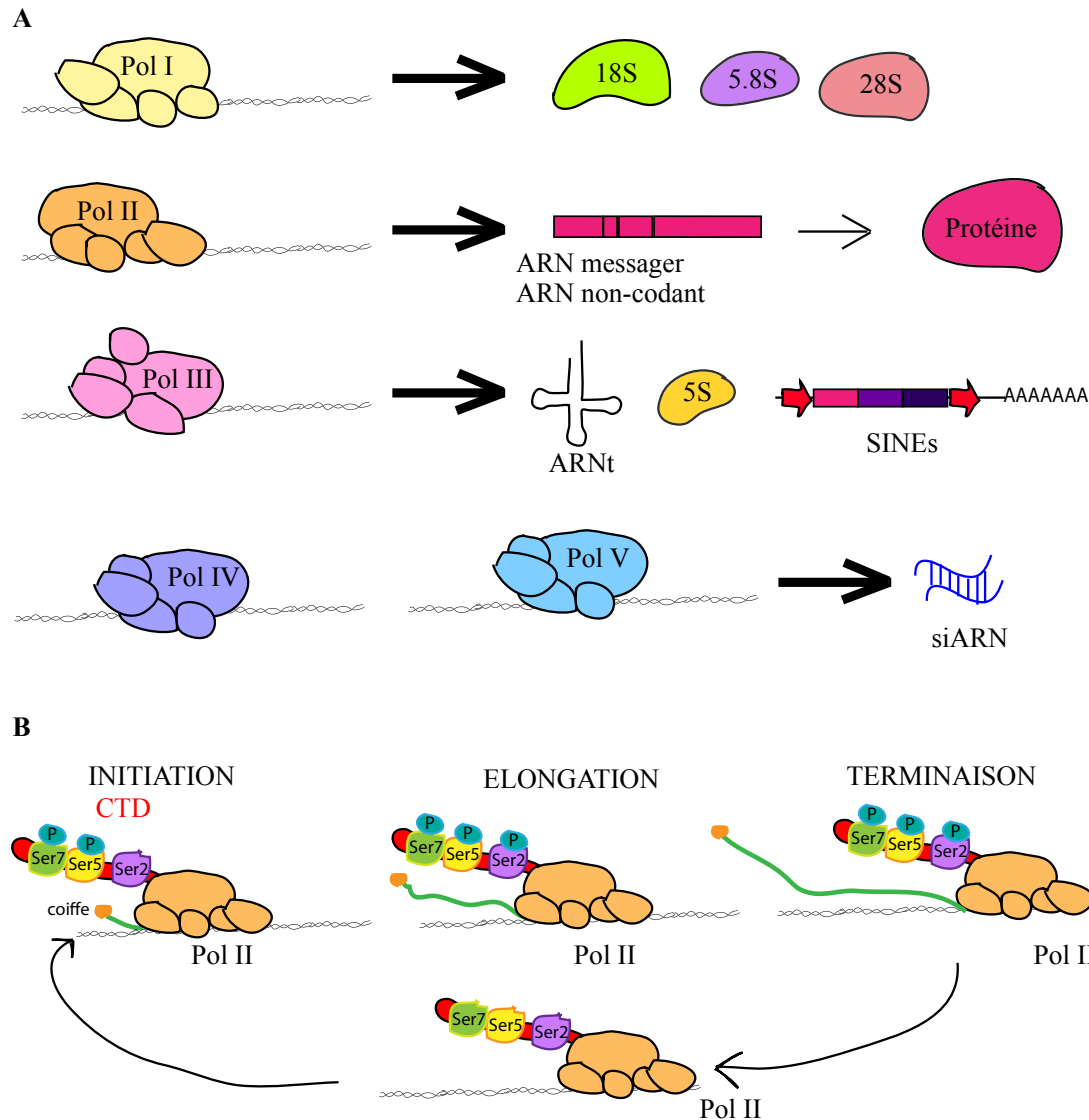
Deux autres polymérases ARN spécifiques des Plantes, Pol IV et Pol V, jouent un rôle crucial dans la génération de petits ARN guidant la méthylation ADN *de novo*, dans un processus appelé voie RNA-directed DNA methylation (RdDM) (Haag et al. 2011).

La transcription par la polymérase II, nécessaire à l'expression des gènes et de la majorité des éléments transposables se déroule en trois phases, l'initiation, l'élongation et la terminaison. Ces étapes sont régulées grâce à des modifications post-traductionnelles du domaine Carboxy-Terminal (CTD) de la Pol II mais également grâce à l'intervention de divers acteurs épigénétiques, comme décrit dans les paragraphes suivants (**Figure I-4B**).

Tout d'abord, le complexe de préinitiation de la transcription recrute la Pol II hypophosphorylée au niveau du promoteur. Le résidu Serine 5 du CTD est phosphorylé par le facteur de transcription TFIIF. Ceci facilite la coiffe en 5' de l'ARNm nouvellement synthétisé. La Serine 7 est également phosphorylée par la cycline kinase CDKF;1 (Hajheidari et al. 2013).

Durant l'élongation, la phosphorylation du CTD sur la Serine 2 par les CDKC ;1 et CDKC ;2 permet aussi son interaction avec de nombreuses protéines impliquées dans l'épissage des transcrits primaires et ainsi le couplage de la transcription et l'épissage.

Chez *Arabidopsis*, les mécanismes de terminaison de la transcription sont encore peu clairs mais le CTD joue également un grand rôle dans le recrutement de la machinerie de maturation de l'extrémité 3' des ARN. Après le clivage et la polyadénylation des ARN, la polymérase est rapidement déphosphorylée afin de pouvoir être recyclée pour un nouveau cycle de transcription.



**Figure I-4 : La transcription chez *Arabidopsis***

**A-** Il existe quatre polymérases chez *A. thaliana*. Tandis que la polymérase II transcrit principalement les gènes codant des protéines ainsi que des ARN non-codants (ARNnc), la polymérase I transcrit les gènes codant les ARN ribosomiques (ARNr) 18S, 5.8 S et 28S. La polymérase III, quant à elle, transcrit les gènes codant les ARN de transfert (ARNt), l'ARNr 5S ainsi que les éléments transposables de classe I de type Short Interspersed Nuclear Elements (SINE). Les polymérases IV et V spécifiques des Plantes permettent la production de petits ARN guidant la méthylation ADN *de novo* via une voie nommée RNA-directed DNA Methylation (RdDM).

**B-** La transcription par la polymérase II se déroule en trois phases, l'initiation où le carboxy-terminal (CTD) est phosphorylé sur les Serines 5 et 7. La phosphorylation facilite le dépôt d'une coiffe sur le transcrit initié. Durant l'élongation, la Serine 2 est phosphorylée également, permettant le couplage de la transcription avec l'épissage. Lors de la terminaison, des protéines de maturation de l'extrémité 3' sont recrutées. Après clivage et polyadénylation du transcrit généré, le CTD est déphosphorylé, la polymérase peut débuter un nouveau cycle de transcription.

#### **I.A.4. Répression des éléments transposables**

Les mécanismes épigénétiques de répression des éléments transposables sont détaillés dans la revue suivante :

« A "mille-feuille" of silencing : epigenetic control of transposable elements. » Rigal M, Mathieu O. ; Biochimica et Biophysica Acta. 2011.

Pour cette revue, j'ai participé à l'étude bibliographique, à la réflexion autour de l'organisation et du contenu de la publication, ainsi qu'à la rédaction des différents paragraphes.





Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: [www.elsevier.com/locate/bbagrm](http://www.elsevier.com/locate/bbagrm)

## Review

A “mille-feuille” of silencing: Epigenetic control of transposable elements<sup>☆</sup>Mélanie Rigal, Olivier Mathieu<sup>\*</sup>

Centre National de la Recherche Scientifique (CNRS), UMR 6247–GrED–INSERM U 931, Clermont Université, Aubière, France

## ARTICLE INFO

## Article history:

Received 15 February 2011

Received in revised form 30 March 2011

Accepted 1 April 2011

Available online 12 April 2011

## Keywords:

Transposable elements

Silencing

Arabidopsis

## ABSTRACT

Despite their abundance in the genome, transposable elements (TEs) and their derivatives are major targets of epigenetic silencing mechanisms, which restrain TE mobility at different stages of the life cycle. DNA methylation, post-translational modification of histone tails and small RNA-based pathways contribute to maintain TE silencing; however, some of these epigenetic marks are tightly interwoven and this complicates the delineation of the exact contribution of each in TE silencing. Recent studies have confirmed that host genomes have evolved versatility in the use of these mechanisms to individualize silencing of particular TEs. These studies also revealed that silencing of TEs is much more dynamic than had been previously thought and can be reversed on the genomic scale in particular cell types or under special environmental conditions. This article is part of a Special Issue entitled “Epigenetic control of cellular and developmental processes in plants”.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

All genomes hold a genetic toolkit that is necessary to properly orchestrate the development of an organism. Genes can be crowded within genomes. Over the past 15 years, intensive sequencing efforts have revealed that most eukaryotic genomes do not simply contain genes but are filled with transposable elements (TEs). In fact, TEs can be the major constituents of a given genome, which leaves little space for genes; this finding is particularly striking in plants [1,2]. TEs are mobile pieces of DNA that have been viewed as genomic parasites that multiply and insert themselves at new genomic positions when activated. TEs have been classified into two classes based on whether their transposition strategy involves an RNA intermediate [3]. Class I TEs, known as retrotransposons, require reverse transcription of an RNA intermediate for duplication, while transposition of class II TEs, also called DNA transposons, does not rely on reverse-transcription of the TE transcripts. DNA transposons are further subdivided into two subclasses. Subclass I is composed of elements containing terminal inverted repeats and are transposed through a “classical” cut-and-paste mechanism. Subclass II contains the more recently identified helitrons, which do not contain terminal inverted repeats and appear to duplicate through a rolling-circle mechanism, which requires the TE-encoded helicase and replicase proteins. Class I elements generally represent the vast majority of the repetitive fraction of a given plant genome, likely due to their duplicative mode of transposition, which increases their copy number after each mobilization.

Although most TEs are present in the form of truncated or mutated derivatives, genomes contain full-length TEs that retain an intact code

and the potential to transpose. Mobilization of these elements represents a threat to genome integrity, a fact that led to their discovery in maize by Barbara McClintock in the late 1940s [4]. Indeed, active TEs can be responsible for chromosome breakage and genomic rearrangements and can interrupt gene function through insertion. To restrain the harmful potential of TEs, host genomes have generated sophisticated mechanisms that counteract TE activation and maintain TEs in a silent state. Although they seem to be selfish pieces of DNA, TEs are thought to play a major role in genome evolution. TEs and the mechanisms that regulate their activity are involved in major cellular processes, such as gene regulation, centromere and telomere function, genomic imprinting and X chromosome inactivation [5,6]. These mechanisms are epigenetic in nature because they do not result from genetic mutations. Rather, they generate a repressive chromatin environment, which is faithfully inherited through mitosis and meiosis, but which can also be reversed in a developmentally regulated or environmentally induced manner. Repressive chromatin is commonly associated with specific small RNA signatures and with a range of epigenetic marks that affect DNA and histone proteins.

In this review, we discuss how these epigenetic processes contribute to the silencing of TEs and focus on recent studies that have expanded our understanding of these pathways.

## 2. Control of TE activity by DNA methylation

DNA methylation, which is the addition of a methyl group to DNA residues, occurs primarily at cytosines in eukaryotes. Apart from yeast (*Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*) and some invertebrates (*Caenorhabditis elegans*, *Drosophila melanogaster*, *Tribolium castaneum*, and *Oikopleura dioica*), cytosine DNA methylation is present at detectable levels in all organisms that have been examined to date [7,8]. In mammals, the prevalent view is that cytosine DNA methylation occurs

<sup>☆</sup> This article is part of a Special Issue entitled “Epigenetic control of cellular and developmental processes in plants”.

<sup>\*</sup> Corresponding author.

E-mail address: [olivier.mathieu@univ-bpclermont.fr](mailto:olivier.mathieu@univ-bpclermont.fr) (O. Mathieu).





almost exclusively at CG dinucleotides, although a recent study has challenged this view by revealing that nearly one-quarter of methylation sites detected in undifferentiated human embryonic stem cells are in non-CG contexts [9]. In plants, non-CG methylation is common, and DNA methylation affects cytosines in all sequence contexts, including CG, CHG (H=A, T or C) and CHH.

Mammalian DNA methyltransferase 1 (DNMT1) and its ortholog in plants, methyltransferase1 (MET1), propagate CG methylation patterns following each round of DNA replication. These DNA methyltransferases function in combination with accessory proteins, UHRF1 in mammals and variant in methylation (VIM) in plants, which have strong affinities for hemimethylated CG sites and likely aid in the recruitment of DNMT1 and MET1 to replicating DNA [10–12]. The plant-specific protein chromomethylase3 (CMT3) primarily maintains CHG methylation, while domains rearranged methyltransferase2 (DRM2) ensures the persistence of CHH methylation. CMT3 is recruited to chromatin via interaction of its chromodomain with methylated histone H3 tails. Accordingly, the CMT3 chromodomain has been shown *in vitro* to bind directly to H3 tails trimethylated at lysine 9 and lysine 27 [13]. DRM2, an ortholog of the mammalian DNMT3 methyltransferases, is guided to DNA by 24-nucleotide (nt) small interfering RNAs (siRNAs) in a pathway known as RNA-directed DNA methylation (RdDM) [14–16]. In addition, decrease in DNA methylation1 (DDM1), a SWI2/SNF2-related chromatin-remodeling factor, contributes to DNA methylation in all sequence contexts [17–19], although its precise molecular mode of action is still poorly understood.

The first evidence for a role of DNA methylation in controlling TE activity resulted from studies of class II TEs in maize, approximately 40 years after the discovery of TEs by B. McClintock. Molecular analyses of the *Activator* (*Ac*), *Suppressor-mutator* (*Spm*) and *Mutator* (*Mu*) elements [20–22] revealed that inactivation of these elements was correlated with the methylation of their DNA. In the active state, transposase promoter sequences from the *Ac* and *Spm* elements are hypomethylated, resulting in active transcription. Likewise, DNA methylation at the autonomous *Mu*-family elements *MuDR* is associated with transcriptional silencing [23,24]. Later studies in *Arabidopsis* showed that several classes of TEs become hypomethylated and transcriptionally reactivated in a *ddm1*-induced hypomethylation background [25–28]. These studies further reported transposition of two DNA transposons in *ddm1*: the MULE-like element *AtMu1* and the *CACTA* elements [25,26]. More recently, the characterization of a *ddm1*-induced floral phenotype revealed that DNA transposons and *ATGP3* retrotransposons of the *gypsy* class are mobilized in *ddm1*. Furthermore, hybridizing genomic DNA from both wild-type and *ddm1* lines on high-density tiling array allowed the authors to detect increases in the copy number, and thus transposition, of several families of DNA transposons and retroelements in the mutant background [29]. Mutations in *LSH1*, the mouse ortholog of *DDM1*, also lead to hypomethylation and transcriptional reactivation of TEs [30,31]. These and other experimental data have led to the view that one important role of DNA methylation is to function as a genomic immune system that ensures the silencing of TEs.

Recent high-throughput profiling studies of DNA methylation in *Arabidopsis* have confirmed that, similar to that observed in human and mouse, TEs and other types of repeated sequences represent the most highly methylated sequences in the genome [32–35]. In two large-scale studies that quantified DNA methylation in an additional set of 21 eukaryotic genomes in total, Zemach et al. and Feng et al. showed that this trend can be extended to all analyzed plant types [7,8]. In addition to CG methylation, TEs are enriched in CHG and CHH methylation, which is consistent with their preferential association with H3K9me2 and their active targeting by the RdDM pathway [32,33,36,37]. It may therefore be reasonable to expect that all three types of methylation cooperate or contribute to a certain degree in TE silencing. Accordingly, *ATGP3* elements that are transcribed and mobilized in the CG and non-CG hypomethylated *ddm1* mutant background remain transcriptionally silent in the *met1* and *cmt3*

single mutants [29]. Transcription and transposition of *ATGP3* was only detected in *met1 cmt3* double mutants, which are deficient in both CG and CHG methylation. Although *CACTA* transcripts accumulate to various degrees in the *met1* and *cmt3* single mutants, high-frequency mobility of this TE is achieved only in the *met1 cmt3* double mutant [38]. Together, these observations illustrate that CG and CHG methylation are redundantly required to restrain TE activity at either the transcriptional (*ATGP3*) or transpositional levels (*CACTA*). These data also suggest that different types of TEs are controlled differentially by DNA methylation. Indeed, this effect was recently illustrated by the analysis of a *cop*ia-like retrotransposon named *Evadé* (*EVD*) [39]. Transcription of *EVD* is reactivated upon loss of CG methylation in a *met1* mutant, and *EVD* is progressively mobilized in successive generations of the mutant. In addition to CG methylation, low levels of CHH methylation are found on this TE. Mutations that affect CHH methylation do not release *EVD* silencing; however, in combination with *met1*, they lead to a synergistic increase in *EVD* transcription, indicating that in the absence of CG methylation, the RdDM pathway curbs the accumulation of *EVD* transcripts. Importantly, in one combination of mutations (*met1 nrpe2*, see below), *EVD* is the only TE mobilized, and other TEs, including *CACTA* and *ATGP3*, remain immobile. These findings underscore the idea that host genomes have evolved selectivity in the control of TE activity.

Unlike mammals, it has long been known that plant genomes likely do not globally reset the genomic DNA methylation patterns between sexual generations. Accordingly, loci that lose DNA methylation and the associated epigenetic silencing in *met1* and *ddm1* mutants, including TEs, retain activity and hypomethylation for several generations following removal of the mutations [17,27,40,41]. Epigenetic recombinant inbred lines (epiRILs) derived from *met1* or *ddm1* mutants show inheritance of DNA hypomethylation at some genomic regions. As expected, inheritance of hypomethylation at *CACTA* elements was found to result in transposition in *ddm1*-derived epiRILs that have inherited these hypomethylated *CACTA* epigenetic alleles (“epialleles”) [42]. In the case of *met1*-derived epiRILs, *CACTA* elements should remain immobile because they do not transpose in the parental *met1* mutant background. Thus, it was unexpected when Reinders et al. reported *CACTA* mobilization in nearly 30% of the *met1*-derived epiRILs. *CACTA* transposition was triggered stochastically in some epiRIL individuals after several generations of propagation by self-pollination [42,43]. Although it was not directly tested, it can be inferred that *CACTA* mobilization in these lines results from the inheritance of *met1*-derived elements that are hypomethylated at CG sites and a concomitant loss of additional repressive epigenetic marks, such as non-CG methylation or H3K9me2. Indeed, it was previously shown that non-CG methylation patterns and H3K9me2 patterns are stochastically altered and unstably inherited upon *met1*-induced loss of CG methylation [44]. Similar to *CACTA*, the *EVD* retroelement was found to transpose after inbreeding of some *met1*-derived epiRILs, but the two types of elements were not necessarily simultaneously reactivated in the same lines, which once again underscores the apparent selective control of TEs by distinct epigenetic mechanisms. Investigating *ddm1*-derived epiRILs, Teixeira and colleagues also reported insights into this selectivity. They demonstrated that a subset of TEs, which is demethylated in *ddm1*, is progressively remethylated and resiled over multiple generations after introduction into a wild-type background [45]. The propensity for remethylation correlates with the retention of CHH methylation, which appears to be independent of DDM1 at remethylatable TEs. In contrast, CHH methylation at non-remethylatable TEs displays a strong dependence on DDM1. Therefore, CHH methylation is clearly maintained by distinct mechanisms at various TEs.

### 3. Role of histone modifications

In addition to DNA methylation, TEs are associated with various post-translational histone modifications, which are characteristic of a



repressive chromatin environment. In particular, TEs located in pericentric heterochromatin are associated with very high levels of H3K9me2 and H3K27me1 [13,36,46,47]. In *Arabidopsis*, several histone methyltransferases are responsible for the propagation of H3K9me2, including kryptonite/SUVH4 (KYP), SUVH5 and SUVH6, which display functional redundancy at some genomic loci [48–52]. Alternatively, H3K27me1 is catalyzed by the histone methyltransferases *Arabidopsis* trithorax-related protein 5 (ATXR5) and ATXR6, which appear to be functionally redundant [46].

H3K9me2 and H3K27me1 both appear to contribute to transcriptional silencing of some TEs. In *Arabidopsis*, TE reactivation was observed in *kyp suvh5 suvh6* triple mutants [49] and *atr5 atr6* double mutants [46]. In rice, mutations in the histone H3K9 methyltransferases genes *SGD714* and *SDG728* also lead to TE activation [53,54]. H3K27me1 functions in transcriptional silencing in a pathway that is clearly independent of DNA methylation and H3K9me2 [46], and TEs reactivated in *atr5 atr6* mutants retain high levels of these two repressive marks. Moreover, H3K27me1 levels are not changed in *met1* and *ddm1* mutants, which show reduced DNA methylation, and the *kyp* mutant, which shows reduced H3K9me2 levels [13,47]. Recently, it was shown that the ATXR5/6 H3K27me1 pathway suppresses re-replication of TE-rich heterochromatic regions, which has led to the speculation that this pathway may have evolved to restrain excess DNA replication and reactivation of TEs [55].

The precise contribution of H3K9me2 to TE silencing is more difficult to dissect given that H3K9me2 is tightly interwoven with DNA methylation, in particular CHG methylation. At the genome level, approximately 90% of CHG methylation overlaps with H3K9me2-enriched regions [36], and TEs have high levels of CHG methylation [28,33,35,56,57]. *KYP* and *KYP SUVH5 SUVH6* mutants show reduced H3K9me2 levels at reactivated TEs and a significant reduction in CHG methylation at these sequences [48–52]. Moreover, some TEs show reduced H3K9me2 levels in mutants for *CMT3*, the maintenance DNA methyltransferase for CHG sites [58]. *KYP*, *SUVH5* and *SUVH6* contain SET and RING associated (SRA) domains, which have been shown to bind to DNA with methylated cytosines, and *KYP* and *SUVH6* display preference for CHG methylation over CG methylation [51,59]. Alternatively, it has been proposed that *CMT3* is recruited to chromatin via the interaction of its chromodomain with H3K9me2 [60]. These findings suggest that H3K9me2 and CHG methylation are maintained through a self-reinforcing feedback loop, such that influencing one modification likely impacts the other. In addition, mutations in *MET1* also reduce H3K9me2 in TE-rich regions [41,61], particularly for TEs with a high frequency of methylated CGs, where H3K9me2 is less dependent upon *KYP* and *CMT3* [62]. This result suggests that an additional pathway, which is dependent on CG methylation, maintains H3K9me2 at TEs. Remarkably, a null mutation in the *increase in bonsai methylation 1* (*IBM1*) histone demethylase gene results in ectopic H3K9me2 and CHG DNA hypermethylation in a large number of genes, while TEs are completely unaffected [62–64]. Thus, *IBM1* protects genes but not TEs from H3K9me2 and CHG methylation. Together with multiple pathways that maintain H3K9me2, this situation contributes to the perpetuation of a robust silenced state at TEs.

Interestingly, it has recently been demonstrated that *KYP* plays a role in restraining the mobilization of the *EVD* retroelement, specifically at the post-transcriptional level [39]. Indeed, while *met1 kyp* double mutants accumulate *EVD* transcripts at a level similar to *met1* single mutants, extrachromosomal copies of this TE only accumulate in the double mutant. In an analogous manner to mammalian lysine methyltransferases, which also methylate non-histone proteins [65–68], it has been proposed that *KYP* may use a substrate other than H3K9 to inactivate, through methylation, a TE-encoded protein required for the translation and/or reverse transcription of *EVD* transcripts [39].

Other histone modifications likely play a role in TE silencing; however, similarly to H3K9me2, the exact contributions of these modifications are difficult to delineate given the links between DNA

methylation and histone modifications. For instance, histone deacetylation appears to be important for the maintenance of TE silencing. In support of this idea, an analysis of *Arabidopsis* histone deacetylase mutants, such as *hda6*, has revealed that several classes of TEs are transcriptionally activated [27]. Likewise, downregulation of the histone deacetylase gene *OsSRT1* in rice results in upregulation of many genomic loci, primarily TEs [69]. Reactivation of these TEs is also correlated with a reduction in H3K9me2, the appearance of H3K4me2, and occasionally reduction in DNA methylation. Similar to histone acetylation, ubiquitination of histone H2B is required to maintain chromatin in an open state [70]. Histone H2B is monoubiquitinated in *Arabidopsis* (at lysine 143), animals and yeast [70–72]. Mutations in an *Arabidopsis* deubiquitination enzyme, UBP36, release transcriptional silencing of several different classes of TEs. Silent TEs are enriched for non-ubiquitinated H2B, and the deubiquitination of H2B seems to be required for H3K9me2 and subsequent non-CG DNA methylation [71]. Interestingly, this suggests that TEs are associated with active histone modification marks by default and that faithful maintenance of a repressive chromatin state depends on the continuous removal of these active marks.

#### 4. Small RNAs and TE silencing

RNA silencing is widely used by eukaryotes, notably to control TE activity at the transcriptional and post-transcriptional levels. This mechanism involves small RNA molecules of 21- to 35-nt, which guide silencing by modulating chromatin states or targeting RNA degradation. Among the various classes of small RNAs, endogenous siRNAs (endo-siRNAs) and PIWI-associated RNAs (piRNAs) participate in the silencing of TEs in both *Drosophila* and mice [73–75]. Endo-siRNAs originate from the cleavage of long double-stranded RNAs (dsRNAs) by a Dicer enzyme; whereas piRNAs are derived from single-stranded transcripts in a Dicer-independent pathway.

The piRNA pathway is mostly germline-specific while endo-siRNAs also repress TE activity in somatic tissues. The endo-siRNA and piRNA pathways have been shown to target heterochromatin formation and transcriptional silencing of TEs in *Drosophila* [76–78]. In mouse, there is also evidence that piRNAs are involved in transcriptional silencing of TEs through DNA methylation [79].

No piRNAs have been reported in plants, but there is abundant evidence that siRNAs are involved in RNA-directed DNA methylation and gene silencing [14]. Genome-wide studies in *Arabidopsis* have shown that siRNAs derived predominantly from TEs and other types of repeats are associated with dense DNA methylation and H3K9me2 [32,33,35,36,80].

The number of proteins identified from genetic screens as being involved in the RdDM pathway is continuously increasing. Proteins required for RdDM include RNA-dependent RNA polymerase2 (RDR2), which generates dsRNAs from single-stranded RNA matrix; the endonuclease DICER-LIKE3, which cleaves dsRNAs to produce 24-nt siRNAs; and Argonaute proteins, including AGO4 and AGO6, which bind siRNAs and play a role in recruiting DNA methyltransferase DRM2 to DNA. Paradoxically, transcription from silent genomic targets is necessary for RdDM. This transcription is achieved by two plant-specific RNA polymerases, Pol IV and Pol V. Pol IV is required for production of the vast majority of 24-nt siRNAs from methylated regions [81], and although Pol IV transcripts have not been identified, they are thought to be used as templates for generation of siRNAs through the successive actions of RDR2 and DCL3. Pol V transcribes intergenic regions, and Pol V transcripts interact with the AGO4/siRNA complexes to induce DNA methylation and H3K9me2 [82,83].

Unlike the *met1* or *ddm1* mutants, which have drastic impacts on TE silencing, mutations in the RdDM pathway in *Arabidopsis* only result in selective reactivation of transcription at a subset of TEs [84–92]. Moreover, TEs are not heritably reactivated by mutations in the RdDM pathway, and the addition of the *met1* mutation to RdDM mutants shows synergistic release of transcriptional silencing [39,44]. This



finding underscores the idea that CG methylation is the primary epigenetic mark of heritable silencing, whereas RdDM-mediated non-CG methylation seems to be largely used to reinforce the pre-existing silencing at TEs. Although the molecular mechanism still needs to be elucidated, it has been recently shown that proteins involved in the RdDM pathway, such as NRPE2 (a common subunit of Pol IV and Pol V), also restrict the mobilization of the *EVD* retroelement at the post-transcriptional level after it has been transcriptionally activated [39].

Initiation of silencing at TEs is still poorly documented; however a role for siRNAs in the establishment of heritable silencing at a TE was clearly demonstrated through studies of the *Mu* killer (*Muk*) locus in maize. This locus naturally occurred as a result of the duplication and inversion of a portion of the autonomous *Mu* element *MuDR* [93]. Transcription of *Muk* produces long RNAs that fold into dsRNAs, thereby providing a template for Dicer-like activities to generate *MuDR*-matching siRNAs [93,94]. When a plant carrying *muk* is crossed with a plant carrying *MuDR*, *Muk*-derived ~25-nt siRNAs induce cytosine DNA methylation in all three sequence contexts at corresponding *MuDR* sequences, associated with an enrichment in H3K9me2 and transcriptional inactivation of *MuDR* [93–95].

Uniquely matching 24-nt siRNAs, which are complementary to a single location in the genome, are more consistently correlated with DNA methylation than siRNAs that match multiple genomic locations [32]. Accordingly, transcription from TEs targeted by uniquely matching siRNAs is more efficiently silenced than TEs targeted by siRNAs that match multiple locations [96]. Hollister and colleagues have suggested that this difference in the efficacy of siRNA-mediated silencing may be explained by the dilution effect of siRNAs that match multiple locations across many targets relative to uniquely matching siRNAs. TEs in *Arabidopsis lyrata*, a close relative of *Arabidopsis thaliana*, are targeted by a lower fraction of uniquely matching siRNAs than in *A. thaliana* [96]. In addition, *A. thaliana* contains two- to threefold fewer TEs than *A. lyrata*, raising the interesting possibility that the variable strength of RdDM may contribute to differential TE proliferation among species [96].

Recently, it was shown that one of the ten Argonaute proteins encoded by the *Arabidopsis* genome, AGO9, controls female gamete formation by restricting the specification of a single megaspore mother cell, the female gametophyte precursor [97]. Wild-type plants develop a single megaspore mother cell that undergoes gametogenesis and forms the female gametophyte, whereas AGO9 mutants have multiple differentiated gametic cells that are able to initiate gametogenesis. A similar phenotype was observed in several mutants of the RdDM pathway, including *rdr2* and *dcl3* and double mutants for Pol IV and Pol V. This phenotype, however, was also observed in mutants for *RDR6* and *SGS3*, which are involved in another siRNA pathway (the trans-acting siRNA pathway) [97]. Interestingly, Olmedo-Monfil and colleagues showed that AGO9 preferentially interacts with 24-nt siRNAs derived from TEs and is necessary for the maintenance of TE silencing in female gametes. These results suggest that AGO9 is involved in an unorthodox, and potentially specific, siRNA-silencing pathway that is necessary for maintaining TE silencing in female gametes. AGO9 is also highly expressed in anthers [97], where it may play a similar role.

## 5. Reversing TE silencing

Although TE silencing can be inherited over multiple generations, it is not static. Different stresses have been shown to counteract silencing, and recent reports have suggested that TE silencing is developmentally controlled and reversed in specific *Arabidopsis* cell types.

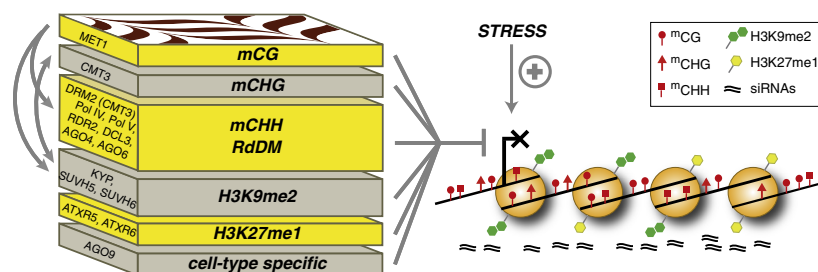
In angiosperms, fertilization involves several nuclei that are contained in both female and male gametophytes. On the female side of the reproductive system, gametogenesis leads to the formation of a haploid egg cell and a homodiploid central cell surrounded by synergid and antipodal haploid cells. The male gametophyte, or pollen grain,

contains three haploid cells: two sperm cells embedded inside the cytoplasm of a larger vegetative cell. Once hydrated on the stigma, this large cell extends a tube that grows through the gynoecium due to attraction by the synergid cells and delivers the two sperm cells inside the female gametophyte. In the double fertilization process specific to angiosperms, one sperm cell fertilizes the haploid egg cell, giving rise to a diploid embryo, while the second sperm cell fertilizes the homodiploid central cell, producing a triploid endosperm that serves as a nurturing tissue for embryo development. Hence, the vegetative cell of the pollen grain and the central cell of the female gametophyte do not contribute to the next generation. Recent reports have shown that drastic changes in DNA methylation and siRNA accumulation take place in these “dead end” cells in *Arabidopsis* [98–101]. Two different technical approaches have revealed widespread reductions in DNA methylation at TEs and other repeats in the endosperm [99,100]. This decreased DNA methylation is in part due to the activity of DEMETER, a DNA glycosylase expressed in the central cell prior to fertilization that excises methylcytosines [99,100,102–104], and also possibly to reduced expression of MET1 during gametogenesis [105]. Hypomethylation occurs primarily at CG sites and is accompanied by increased methylation at CHH sites in repeated sequences, which are consistent with previous observations from *met1* mutants that have lost CG methylation [44,99]. Methylation at CHH sites is a signature of RdDM, and consistent with this idea, Mosher and colleagues have reported a massive burst of 24-nt siRNAs from TEs and repeats. This burst begins in the central cell and continues in the endosperm during seed development [98]. Whether reduced DNA methylation results in the release of TE silencing in the endosperm has not been tested. Slotkin and colleagues, however, have shown transcriptional reactivation and mobilization of some TEs that are associated with reduced DNA methylation and a loss of 24-nt TE siRNAs in the vegetative cell of the pollen grain [101]. Hypomethylation in the vegetative nucleus is likely a consequence of both downregulation of genes required for the maintenance of DNA methylation, such as *MET1* and *DDM1*, and of active demethylation by an unidentified DNA demethylase [101,105,106]. Slotkin et al. also noticed an accumulation of 21-nt siRNAs from *Athila* retrotransposons, which may be produced in the vegetative nucleus and accumulate in sperm cells [101]. Similarly, because both 21- and 24-nt siRNAs are mobile in *Arabidopsis* [107,108], the burst of 24-nt siRNAs produced in the central cell/endosperm may migrate into and accumulate in the egg cell/embryo. This high level of 24-nt siRNAs may provide an additional layer of silencing at TEs in the egg cell/embryo by reinforcing the RdDM pathway. Consistent with this explanation, Hsieh and colleagues found that the levels of CHH methylation at TEs and repeats are elevated in the embryo relative to adult tissues [99]. On the male side, 21-nt siRNAs produced in the vegetative cell may also strengthen TE silencing in sperm cells at the post-transcriptional level to prevent TE mobilization. These 21- and 24-nt siRNAs may play similar roles in backup of TE silencing during fertilization [78].

Epigenetic reprogramming of TE silencing has also been documented in maize, where *MuDR* silencing is reversed upon the change from juvenile to adult phase, in a tissue adjacent to the one that will produce the germ line [95].

Plants do not establish a germ cell line early during development. Instead, progenitor germ cells differentiate late during sporophyte development. The optimum window for a TE to mobilize and produce new insertions that will be subsequently inherited by the next generation is between the differentiation of gametophyte precursors and the formation of the early embryo. In this regard, accumulating extra layers of TE silencing in these cells would secure genome integrity. But, what could induce TE silencing relaxation? Obviously, as demonstrated experimentally, mutations in epigenetic regulators involved in the maintenance of CG methylation, such as *MET1* or *DDM1*, may induce TE mobilization [25,26,29,39]. However, such mutations would be expected to occur relatively rarely. Perhaps more





**Fig. 1.** Representation of the different layers that constitute the “mille-feuille” of silencing (left), which ensures efficient silencing of TEs (right). Repression of TEs depends on the combined contribution of the different layers of silencing, and the importance of each layer is variable across TEs. In addition, crosstalk exists between some of the epigenetic marks that constitute these silencing layers (arrows on the left). Environmental stresses can overcome known repressive epigenetic marks and lead to reactivation of some TEs.

commonly, changes in environmental conditions or sudden environmental stresses could challenge TE silencing. Accordingly, two recent reports have demonstrated that heat stress can overcome TE silencing in *Arabidopsis*, at least at the transcriptional level [109,110]. Transcription of several classes of TEs and other heterochromatic targets, which are normally silent at ambient temperature, is significantly upregulated upon exposure to a heat stress at 37 °C. This release from silencing occurs across the genome [109] and is mainly transient. Interestingly, at the loci examined, stress-induced destabilization of silencing occurred without altering common repressive epigenetic marks, including DNA methylation and H3K9me2 [109,110], indicating that these marks are not sufficient for efficient transcriptional silencing under these conditions. It is currently unknown whether heat stress produces an activating signal that overcomes the presence of these marks or interferes with their readout. Moreover, whether stresses have the same effect on the genomes of gametes, gamete precursors or the early embryo is not known. It would be of interest to determine if the “extra layer” of silencing present in these cell types/developmental stages, which are represented by high levels of specific siRNAs and the possible action of an unconventional siRNA pathway involving AGO9, provides these cells/developmental stages with relatively lower sensitivity to stress.

In a recent study, Ito and colleagues showed that an LTR retroelement, named *ONSEN*, could be mobilized following heat stress-induced transcriptional stimulation [111]. Although stress was applied to young seedlings, transposition of *ONSEN* was found only in the progeny of stressed plants. Furthermore, new *ONSEN* insertions were detected only in the progeny of RdDM mutants subjected to stress but not in the progeny of stressed wild-type individuals. Using elegant genetic approaches, the authors demonstrated that in the absence of a functional RdDM pathway, “memory” of the heat-stress persists during plant development, resulting in *ONSEN* mobilization, which interestingly occurred prior to gametophyte development [111].

## 6. Conclusion

A combination of genomic and genetic studies in plants have revealed that TEs are the main targets of epigenetic silencing pathways and that TEs are controlled by several layers of silencing. Crosstalk among these layers has been observed, and as exemplified by specific bursts of 21- and 24-nt siRNAs and the role of AGO9, certain silencing layers seem to be devoted to specific cell types or particular developmental stages. In this “mille-feuille” of silencing, the precise contribution of each layer to the silencing efficiency for a given TE is highly variable (Fig. 1).

The purpose and evolutionary origin for this specificity remain elusive. Because TEs represent highly dynamic genomic components and because there are potentially deleterious consequences of TE mobilization, evolutionary forces have likely driven the production of

various epigenetic mechanisms to ensure TE silencing. Various factors may govern the specificity of TE silencing, including TE size, copy number, transposition competence, genomic location and local chromatin environment. Accumulating evidence has demonstrated that a variety of environmental stresses can challenge TE silencing and possibly lead to bursts of transposition. Interestingly, distinct stresses affect TEs differentially, suggesting that selectivity and complexity in TE silencing may have evolved as a consequence of stress exposure.

As we increase our knowledge of the molecular details of TE silencing, future studies will need to reveal how environmental factors interfere with these silencing mechanisms and thereby contribute to TE proliferation.

## Acknowledgments

We thank Thierry Pélissier and Charles White for comments on the article and the anonymous reviewers for helpful suggestions. We apologize to colleagues whose research we did not have space to discuss. O.M. is supported by a Starting Independent Researcher grant from the European Research Council (260742-I2ST).

## References

- [1] M.I. Tenaillon, J.D. Hollister, B.S. Gaut, A triptych of the evolution of plant transposable elements, *Trends Plant Sci.* 15 (2010) 471–478.
- [2] S.R. Wessler, Transposable elements and the evolution of eukaryotic genomes, *Proc. Natl. Acad. Sci. USA* 103 (2006) 17600–17601.
- [3] T. Wicker, F. Sabot, A. Hua-Van, J.L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, A.H. Schulman, A unified classification system for eukaryotic transposable elements, *Nat. Rev. Genet.* 8 (2007) 973–982.
- [4] B. McClintock, Mutable loci in maize, *Carnegie Inst. Wash. Year b* 47 (1948) 155–169.
- [5] D. Lisch, Epigenetic regulation of transposable elements in plants, *Annu. Rev. Plant Biol.* 60 (2009) 43–66.
- [6] R.K. Slotkin, R. Martienssen, Transposable elements and the epigenetic regulation of the genome, *Nat. Rev. Genet.* 8 (2007) 272–285.
- [7] S. Feng, S.J. Cokus, X. Zhang, P.Y. Chen, M. Bostick, M.G. Goll, J. Hetzel, J. Jain, S.H. Strauss, M.E. Halpern, C. Ukomadu, K.C. Sadler, S. Pradhan, M. Pellegrini, S.E. Jacobsen, Conservation and divergence of methylation patterning in plants and animals, *Proc. Natl. Acad. Sci. USA* 107 (2010) 8689–8694.
- [8] A. Zemach, I.E. McDaniel, P. Silva, D. Zilberman, Genome-wide evolutionary analysis of eukaryotic DNA methylation, *Science* 328 (2010) 916–919.
- [9] R. Lister, M. Pelizzola, R.H. Dowen, R.D. Hawkins, G. Hon, J. Tonti-Filippini, J.R. Nery, L. Lee, Z. Ye, Q.M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A.H. Millar, J.A. Thomson, B. Ren, J.R. Ecker, Human DNA methylomes at base resolution show widespread epigenomic differences, *Nature* 462 (2009) 315–322.
- [10] M. Bostick, J.K. Kim, P.O. Esteve, A. Clark, S. Pradhan, S.E. Jacobsen, UHRF1 plays a role in maintaining DNA methylation in mammalian cells, *Science* 317 (2007) 1760–1764.
- [11] S. Liu, Y. Yu, Y. Ruan, D. Meyer, M. Wolff, L. Xu, N. Wang, A. Steinmetz, W.H. Shen, Plant SET- and RING-associated domain proteins in heterochromatinization, *Plant J.* 52 (2007) 914–926.
- [12] H.R. Woo, T.A. Dittmer, E.J. Richards, Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in *Arabidopsis*, *PLoS Genet.* 4 (2008) e1000156.

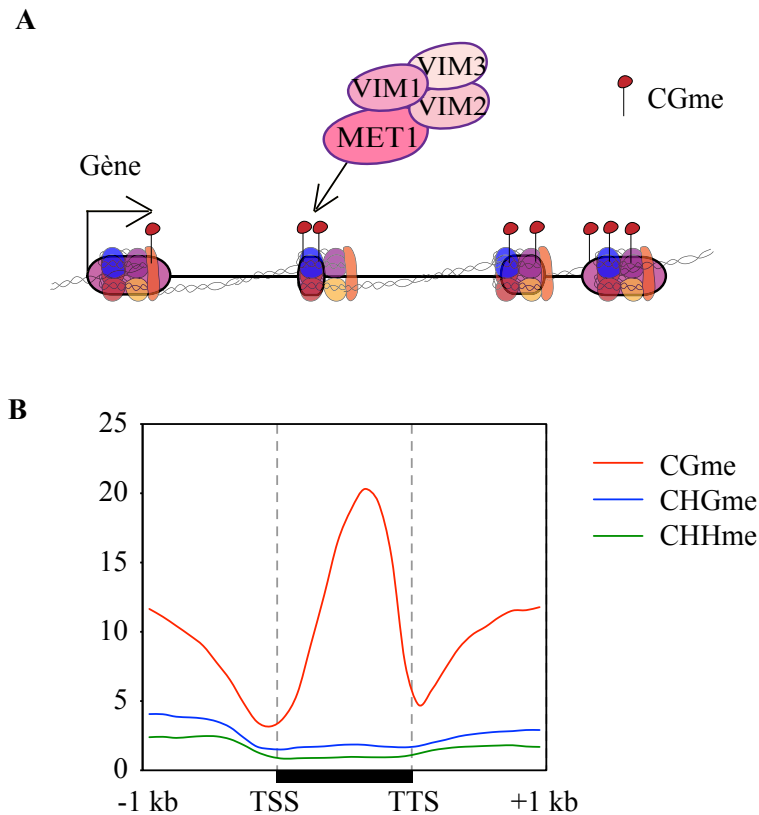




- [13] A.M. Lindroth, D. Shultz, Z. Jasencakova, J. Fuchs, L. Johnson, D. Schubert, D. Patnaik, S. Pradhan, J. Goodrich, I. Schubert, T. Jenuwein, S. Khorasanizadeh, S.E. Jacobsen, Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3, *EMBO J.* 23 (2004) 4286–4296.
- [14] M. Matzke, T. Kanno, L. Daxinger, B. Huettel, A.J. Matzke, RNA-mediated chromatin-based silencing in plants, *Curr. Opin. Cell Biol.* 21 (2009) 367–376.
- [15] X. Cao, S.E. Jacobsen, Role of the *Arabidopsis* DRM methyltransferases in de novo DNA methylation and gene silencing, *Curr. Biol.* 12 (2002) 1138–1144.
- [16] X. Cao, W. Aufsatz, D. Zilberman, M.F. Mette, M.S. Huang, M. Matzke, S.E. Jacobsen, Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation, *Curr. Biol.* 13 (2003) 2212–2217.
- [17] A. Vongs, T. Kakutani, R.A. Martienssen, E.J. Richards, *Arabidopsis thaliana* DNA methylation mutants, *Science* 260 (1993) 1926–1928.
- [18] J.A. Jeddeloh, T.L. Stokes, E.J. Richards, Maintenance of genomic methylation requires a SWI2/SNF2-like protein, *Nat. Genet.* 22 (1999) 94–97.
- [19] J.A. Jeddeloh, J. Bender, E.J. Richards, The DNA methylation locus DDM1 is required for maintenance of gene silencing in *Arabidopsis*, *Genes Dev.* 12 (1998) 1714–1725.
- [20] J.A. Banks, P. Masson, N. Fedoroff, Molecular mechanisms in the developmental regulation of the maize suppressor-mutator transposable element, *Genes Dev.* 2 (1988) 1364–1380.
- [21] V.L. Chandler, V. Walbot, DNA modification of a maize transposable element correlates with loss of activity, *Proc. Natl. Acad. Sci. USA* 83 (1986) 1767–1771.
- [22] P.S. Chomet, S. Wessler, S.L. Dellaporta, Inactivation of the maize transposable element Activator (Ac) is associated with its DNA modification, *EMBO J.* 6 (1987) 295–302.
- [23] R.J. Hershberger, C.A. Warren, V. Walbot, Mutator activity in maize correlates with the presence and expression of the Mu transposable element Mu9, *Proc. Natl. Acad. Sci. USA* 88 (1991) 10198–10202.
- [24] G.N. Rudenko, V. Walbot, Expression and post-transcriptional regulation of maize transposable element MuDR and its derivatives, *Plant Cell* 13 (2001) 553–570.
- [25] A. Miura, S. Yonebayashi, K. Watanabe, T. Toyama, H. Shimada, T. Kakutani, Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*, *Nature* 411 (2001) 212–214.
- [26] T. Singer, C. Yordan, R.A. Martienssen, Robertson's mutator transposons in *A. thaliana* are regulated by the chromatin-remodeling gene decrease in DNA methylation (DDM1), *Genes Dev.* 15 (2001) 591–602.
- [27] Z. Lippman, B. May, C. Yordan, T. Singer, R. Martienssen, Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification, *PLoS Biol.* 1 (2003) E67.
- [28] Z. Lippman, A.V. Gendrel, M. Black, M.W. Vaughn, N. Dedhia, W.R. McCombie, K. Lavine, V. Mittal, B. May, K.D. Kasschau, J.C. Carrington, R.W. Doerge, V. Colot, R. Martienssen, Role of transposable elements in heterochromatin and epigenetic control, *Nature* 430 (2004) 471–476.
- [29] S. Tsukahara, A. Kobayashi, A. Kawabe, O. Mathieu, A. Miura, T. Kakutani, Bursts of retrotransposition reproduced in *Arabidopsis*, *Nature* 461 (2009) 423–426.
- [30] Q. Yan, E. Cho, S. Lockett, K. Muegge, Association of Lsh, a regulator of DNA methylation, with pericentromeric heterochromatin is dependent on intact heterochromatin, *Mol. Cell. Biol.* 23 (2003) 8416–8428.
- [31] J. Huang, T. Fan, Q. Yan, H. Zhu, S. Fox, H.J. Issaq, L. Best, L. Gangi, D. Munroe, K. Muegge, Lsh, an epigenetic guardian of repetitive elements, *Nucleic Acids Res.* 32 (2004) 5019–5028.
- [32] R. Lister, R.C. O'Malley, J. Tonti-Filippini, B.D. Gregory, C.C. Berry, A.H. Millar, J.R. Ecker, Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*, *Cell* 133 (2008) 523–536.
- [33] X. Zhang, J. Yazaki, A. Sundaresan, S. Cokus, S.W. Chan, H. Chen, I.R. Henderson, P. Shinn, M. Pellegrini, S.E. Jacobsen, J.R. Ecker, Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*, *Cell* 126 (2006) 1189–1201.
- [34] D. Zilberman, S. Henikoff, Genome-wide analysis of DNA methylation patterns, *Development* 134 (2007) 3959–3965.
- [35] S.J. Cokus, S. Feng, X. Zhang, Z. Chen, B. Merriman, C.D. Haudenschild, S. Pradhan, S.F. Nelson, M. Pellegrini, S.E. Jacobsen, Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning, *Nature* 452 (2008) 215–219.
- [36] Y.V. Bernatavichute, X. Zhang, S. Cokus, M. Pellegrini, S.E. Jacobsen, Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*, *PLoS One* 3 (2008) e3156.
- [37] X. Zhang, I.R. Henderson, C. Lu, P.J. Green, S.E. Jacobsen, Role of RNA polymerase IV in plant small RNA metabolism, *Proc. Natl. Acad. Sci. USA* 104 (2007) 4536–4541.
- [38] M. Kato, A. Miura, J. Bender, S.E. Jacobsen, T. Kakutani, Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*, *Curr. Biol.* 13 (2003) 421–426.
- [39] M. Mirouze, J. Reinders, E. Bucher, T. Nishimura, K. Schneeberger, S. Ossowski, J. Cao, D. Weigel, J. Paszkowski, O. Mathieu, Selective epigenetic control of retrotransposition in *Arabidopsis*, *Nature* 461 (2009) 427–430.
- [40] T. Kakutani, K. Munakata, E.J. Richards, H. Hirochika, Meiotically and mitotically stable inheritance of DNA hypomethylation induced by ddm1 mutation of *Arabidopsis thaliana*, *Genetics* 151 (1999) 831–838.
- [41] W.J. Soppe, Z. Jasencakova, A. Houben, T. Kakutani, A. Meister, M.S. Huang, S.E. Jacobsen, I. Schubert, P.F. Franz, DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in *Arabidopsis*, *EMBO J.* 21 (2002) 6549–6559.
- [42] F. Johannes, E. Porcher, F.K. Teixeira, V. Saliba-Colombani, M. Simon, N. Agier, A. Bulski, J. Albuissou, F. Heredia, P. Audigier, D. Bouchez, C. Dillmann, P. Guerche, F. Hospital, V. Colot, Assessing the impact of transgenerational epigenetic variation on complex traits, *PLoS Genet.* 5 (2009) e1000530.
- [43] J. Reinders, B.B. Wulff, M. Mirouze, A. Mari-Ordóñez, M. Dapp, W. Rozhon, E. Bucher, G. Theiler, J. Paszkowski, Compromised stability of DNA methylation and transposon immobilization in mosaic *Arabidopsis* epigenomes, *Genes Dev.* 23 (2009) 939–950.
- [44] O. Mathieu, J. Reinders, M. Caikovski, C. Smathajitt, J. Paszkowski, Transgenerational stability of the *Arabidopsis* epigenome is coordinated by CG methylation, *Cell* 130 (2007) 851–862.
- [45] F.K. Teixeira, F. Heredia, A. Sarazin, F. Roudier, M. Boccara, C. Claudio, C. Cruaud, J. Poulain, M. Berdasco, M.F. Fraga, O. Voinnet, P. Wincker, M. Esteller, V. Colot, A role for RNAi in the selective correction of DNA methylation defects, *Science* 323 (2009) 1600–1604.
- [46] Y. Jacob, S. Feng, C.A. LeBlanc, Y.V. Bernatavichute, H. Stroud, S. Cokus, L.M. Johnson, M. Pellegrini, S.E. Jacobsen, S.D. Michaels, ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing, *Nat. Struct. Mol. Biol.* 16 (2009) 763–768.
- [47] O. Mathieu, A.V. Probst, J. Paszkowski, Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in *Arabidopsis*, *EMBO J.* 24 (2005) 2783–2791.
- [48] M.L. Ebbs, L. Barte, J. Bender, H3 lysine 9 methylation is maintained on a transcribed inverted repeat by combined action of SUVH6 and SUVH4 methyltransferases, *Mol. Cell. Biol.* 25 (2005) 10507–10515.
- [49] M.L. Ebbs, J. Bender, Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase, *Plant Cell* 18 (2006) 1166–1176.
- [50] J.P. Jackson, A.M. Lindroth, X. Cao, S.E. Jacobsen, Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase, *Nature* 416 (2002) 556–560.
- [51] L.M. Johnson, M. Bostick, X. Zhang, E. Kraft, I. Henderson, J. Callis, S.E. Jacobsen, The SRA methyl-cytosine-binding domain links DNA and histone methylation, *Curr. Biol.* 17 (2007) 379–384.
- [52] F. Malagnac, L. Barte, J. Bender, An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation, *EMBO J.* 21 (2002) 6842–6852.
- [53] Y. Ding, X. Wang, L. Su, J. Zhai, S. Cao, D. Zhang, C. Liu, Y. Bi, Q. Qian, Z. Cheng, C. Chu, X. Cao, SDG714, a histone H3K9 methyltransferase, is involved in Tos17 DNA methylation and transposition in rice, *Plant Cell* 19 (2007) 9–22.
- [54] F.J. Qin, Q.W. Sun, L.M. Huang, X.S. Chen, D.X. Zhou, Rice SUVH histone methyltransferase genes display specific functions in chromatin modification and retrotransposon repression, *Mol. Plant* 3 (2010) 773–782.
- [55] Y. Jacob, H. Stroud, C. LeBlanc, S. Feng, L. Zhuo, E. Caro, C. Hassel, C. Gutierrez, S.D. Michaels, S.E. Jacobsen, Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases, *Nature* 466 (2010) 987–991.
- [56] R. Tompa, C.M. McCallum, J. Delrow, J.G. Henikoff, B. van Steensel, S. Henikoff, Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3, *Curr. Biol.* 12 (2002) 65–68.
- [57] D. Zilberman, M. Gehring, R.K. Tran, T. Ballinger, S. Henikoff, Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription, *Nat. Genet.* 39 (2007) 61–69.
- [58] L. Johnson, X. Cao, S. Jacobsen, Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation, *Curr. Biol.* 12 (2002) 1360–1367.
- [59] E. Rajakumara, J.A. Law, D.K. Simanshu, P. Voigt, L.M. Johnson, D. Reinberg, D.J. Patel, S.E. Jacobsen, A dual flip-out mechanism for 5mC recognition by the *Arabidopsis* SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo, *Genes Dev.* 25 (2011) 137–152.
- [60] S. Feng, S.E. Jacobsen, Epigenetic modifications in plants: an evolutionary perspective, *Current opinion in plant biology*, 2011.
- [61] M. Tariq, H. Saze, A.V. Probst, J. Lichota, Y. Habu, J. Paszkowski, Erasure of CpG methylation in *Arabidopsis* alters patterns of histone H3 methylation in heterochromatin, *Proc. Natl. Acad. Sci. USA* 100 (2003) 8823–8827.
- [62] S. Inagaki, A. Miura-Kamio, Y. Nakamura, F. Lu, X. Cui, X. Cao, H. Kimura, H. Saze, T. Kakutani, Autocatalytic differentiation of epigenetic modifications within the *Arabidopsis* genome, *EMBO J.* 29 (2010) 3496–3506.
- [63] A. Miura, M. Nakamura, S. Inagaki, A. Kobayashi, H. Saze, T. Kakutani, An *Arabidopsis* jmjC domain protein protects transcribed genes from DNA methylation at CHG sites, *EMBO J.* 28 (2009) 1078–1086.
- [64] H. Saze, A. Shiraishi, A. Miura, T. Kakutani, Control of genic DNA methylation by a jmjC domain-containing protein in *Arabidopsis thaliana*, *Science* 319 (2008) 462–465.
- [65] A. Kouskouti, E. Scheer, A. Staub, L. Tora, I. Talianidis, Gene-specific modulation of TAF10 function by SET9-mediated methylation, *Mol. Cell* 14 (2004) 175–182.
- [66] S.C. Sampath, I. Marazzi, K.L. Yap, A.N. Krutchinsky, I. Mecklenbrauer, A. Viale, E. Rudensky, M.M. Zhou, B.T. Chait, A. Tarakhovskiy, Methylation of a histone mimic within the histone methyltransferase G9a regulates protein complex assembly, *Mol. Cell* 27 (2007) 596–608.
- [67] P.O. Esteve, H.G. Chin, J. Benner, G.R. Feehery, M. Samaranyake, G.A. Horwitz, S.E. Jacobsen, S. Pradhan, Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells, *Proc. Natl. Acad. Sci. USA* 106 (2009) 5076–5081.
- [68] S. Chukhov, J.K. Kurash, J.R. Wilson, B. Xiao, N. Justin, G.S. Ivanov, K. McKinney, P. Tempst, C. Prives, S.J. Gamblin, N.A. Barlev, D. Reinberg, Regulation of p53 activity through lysine methylation, *Nature* 432 (2004) 353–360.
- [69] L. Huang, Q. Sun, F. Qin, C. Li, Y. Zhao, D.X. Zhou, Down-regulation of a SILENT INFORMATION REGULATOR2-related histone deacetylase gene, OsSRT1, induces DNA fragmentation and cell death in rice, *Plant Physiol.* 144 (2007) 1508–1519.
- [70] Y. Zhang, Transcriptional regulation by histone ubiquitination and deubiquitination, *Genes Dev.* 17 (2003) 2733–2740.
- [71] V.V. Sridhar, A. Kapoor, K. Zhang, J. Zhu, T. Zhou, P.M. Hasegawa, R.A. Bressan, J.K. Zhu, Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination, *Nature* 447 (2007) 735–738.



- [72] Z.W. Sun, C.D. Allis, Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast, *Nature* 418 (2002) 104–108.
- [73] V.N. Kim, J. Han, M.C. Siomi, Biogenesis of small RNAs in animals, *Nat. Rev. Mol. Cell Biol.* 10 (2009) 126–139.
- [74] K. Okamura, E.C. Lai, Endogenous small interfering RNAs in animals, *Nat. Rev. Mol. Cell Biol.* 9 (2008) 673–678.
- [75] M.C. Siomi, K. Sato, D. Pezic, A.A. Aravin, PIWI-interacting small RNAs: the vanguard of genome defence, *Nat. Rev. Mol. Cell Biol.* 12 (2011) 246–258.
- [76] D. Fagegaltier, A.L. Bouge, B. Berry, E. Poisot, O. Sismeiro, J.Y. Coppee, L. Theodore, O. Voinnet, C. Antoniewski, The endogenous siRNA pathway is involved in heterochromatin formation in *Drosophila*, *Proc. Natl. Acad. Sci. USA* 106 (2009) 21258–21263.
- [77] M. Pal-Bhadra, B.A. Leibovitch, S.G. Gandhi, M. Rao, U. Bhadra, J.A. Birchler, S.C. Elgin, Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery, *Science* 303 (2004) 669–672.
- [78] D. Bourc'his, O. Voinnet, A small-RNA perspective on gametogenesis, fertilization, and early zygotic development, *Science* 330 (2010) 617–622.
- [79] A.A. Aravin, R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K.F. Toth, T. Bestor, G.J. Hannon, A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice, *Mol. Cell* 31 (2008) 785–799.
- [80] K.D. Kasschau, N. Fahlgren, E.J. Chapman, C.M. Sullivan, J.S. Cumbie, S.A. Givan, J.C. Carrington, Genome-wide profiling and analysis of *Arabidopsis* siRNAs, *PLoS Biol.* 5 (2007) e57.
- [81] R.A. Mosher, F. Schwach, D. Studholme, D.C. Baulcombe, PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis, *Proc. Natl. Acad. Sci. USA* 105 (2008) 3145–3150.
- [82] A.T. Wierzbicki, J.R. Haag, C.S. Pikaard, Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes, *Cell* 135 (2008) 635–648.
- [83] A.T. Wierzbicki, T.S. Ream, J.R. Haag, C.S. Pikaard, RNA polymerase V transcription guides ARGONAUTE4 to chromatin, *Nat. Genet.* 41 (2009) 630–634.
- [84] A.J. Herr, M.B. Jensen, T. Dalmay, D.C. Baulcombe, RNA polymerase IV directs silencing of endogenous DNA, *Science* 308 (2005) 118–120.
- [85] T. Kanno, E. Bucher, L. Daxinger, B. Huettel, D.P. Kreil, F. Breinig, M. Lind, M.J. Schmitt, S.A. Simon, S.G. Guazada, B.C. Meyers, Z.J. Lorkovic, A.J. Matzke, M. Matzke, RNA-directed DNA methylation and plant development require an IWR1-type transcription factor, *EMBO Rep.* 11 (2010) 65–71.
- [86] T. Kanno, M.F. Mette, D.P. Kreil, W. Aufsatz, M. Matzke, A.J. Matzke, Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation, *Curr. Biol.* 14 (2004) 801–805.
- [87] T. Kanno, B. Huettel, M.F. Mette, W. Aufsatz, E. Jaligot, L. Daxinger, D.P. Kreil, M. Matzke, A.J. Matzke, Atypical RNA polymerase subunits required for RNA-directed DNA methylation, *Nat. Genet.* 37 (2005) 761–765.
- [88] T. Kanno, E. Bucher, L. Daxinger, B. Huettel, G. Bohmdorfer, W. Gregor, D.P. Kreil, M. Matzke, A.J. Matzke, A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation, *Nat. Genet.* 40 (2008) 670–675.
- [89] Y. Onodera, J.R. Haag, T. Ream, P.C. Nunes, O. Pontes, C.S. Pikaard, Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation, *Cell* 120 (2005) 613–622.
- [90] B. Huettel, T. Kanno, L. Daxinger, W. Aufsatz, A.J. Matzke, M. Matzke, Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*, *EMBO J.* 25 (2006) 2828–2836.
- [91] Z. Gao, H.L. Liu, L. Daxinger, O. Pontes, X. He, W. Qian, H. Lin, M. Xie, Z.J. Lorkovic, S. Zhang, D. Miki, X. Zhan, D. Pontier, T. Lagrange, H. Jin, A.J. Matzke, M. Matzke, C.S. Pikaard, J.K. Zhu, An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation, *Nature* 465 (2010) 106–109.
- [92] D. Pontier, G. Yahubyan, D. Vega, A. Bulski, J. Saez-Vasquez, M.A. Hakimi, S. Lerbs-Mache, V. Colot, T. Lagrange, Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*, *Genes Dev.* 19 (2005) 2030–2040.
- [93] R.K. Slotkin, M. Freeling, D. Lisch, Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication, *Nat. Genet.* 37 (2005) 641–644.
- [94] R.K. Slotkin, M. Freeling, D. Lisch, Mu killer causes the heritable inactivation of the Mutator family of transposable elements in *Zea mays*, *Genetics* 165 (2003) 781–797.
- [95] H. Li, M. Freeling, D. Lisch, Epigenetic reprogramming during vegetative phase change in maize, *Proc. Natl. Acad. Sci. USA* 107 (2010) 22184–22189.
- [96] J.D. Hollister, L.M. Smith, Y.L. Guo, F. Ott, D. Weigel, B.S. Gaut, Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*, *Proc. Natl. Acad. Sci. USA* (2011).
- [97] V. Olmedo-Monfil, N. Duran-Figueroa, M. Arteaga-Vazquez, E. Demesa-Arevalo, D. Autran, D. Grimanelli, R.K. Slotkin, R.A. Martienssen, J.P. Vielle-Calzada, Control of female gamete formation by a small RNA pathway in *Arabidopsis*, *Nature* 464 (2010) 628–632.
- [98] R.A. Mosher, C.W. Melnyk, K.A. Kelly, R.M. Dunn, D.J. Studholme, D.C. Baulcombe, Uniparental expression of PolIV-dependent siRNAs in developing endosperm of *Arabidopsis*, *Nature* 460 (2009) 283–286.
- [99] T.F. Hsieh, C.A. Ibarra, P. Silva, A. Zemach, L. Eshed-Williams, R.L. Fischer, D. Zilberman, Genome-wide demethylation of *Arabidopsis* endosperm, *Science* 324 (2009) 1451–1454.
- [100] M. Gehring, K.L. Bubbs, S. Henikoff, Extensive demethylation of repetitive elements during seed development underlies gene imprinting, *Science* 324 (2009) 1447–1451.
- [101] R.K. Slotkin, M. Vaughn, F. Borges, M. Tanurdzic, J.D. Becker, J.A. Feijo, R.A. Martienssen, Epigenetic reprogramming and small RNA silencing of transposable elements in pollen, *Cell* 136 (2009) 461–472.
- [102] M. Gehring, J.H. Huh, T.F. Hsieh, J. Pennerman, Y. Choi, J.J. Harada, R.B. Goldberg, R.L. Fischer, DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation, *Cell* 124 (2006) 495–506.
- [103] T. Morales-Ruiz, A.P. Ortega-Galisteo, M.I. Ponferrada-Marin, M.I. Martinez-Macias, R.R. Ariza, T. Roldan-Arjona, DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases, *Proc. Natl. Acad. Sci. USA* 103 (2006) 6853–6858.
- [104] Y. Choi, M. Gehring, L. Johnson, M. Hannon, J.J. Harada, R.B. Goldberg, S.E. Jacobsen, R.L. Fischer, DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*, *Cell* 110 (2002) 33–42.
- [105] P.E. Jullien, A. Mosquana, M. Ingouff, T. Sakata, N. Ohad, F. Berger, Retinoblastoma and its binding partner MSI1 control imprinting in *Arabidopsis*, *PLoS Biol.* 6 (2008) e194.
- [106] F. Borges, G. Gomes, R. Gardner, N. Moreno, S. McCormick, J.A. Feijo, J.D. Becker, Comparative transcriptomics of *Arabidopsis* sperm cells, *Plant Physiol.* 148 (2008) 1168–1181.
- [107] P. Dunoyer, C.A. Brosnan, G. Schott, Y. Wang, F. Jay, A. Alioua, C. Himber, O. Voinnet, An endogenous, systemic RNAi pathway in plants, *EMBO J.* 29 (2010) 1699–1712.
- [108] A. Molnar, C.W. Melnyk, A. Bassett, T.J. Hardcastle, R. Dunn, D.C. Baulcombe, Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells, *Science* 328 (2010) 872–875.
- [109] M. Tittel-Elmer, E. Bucher, L. Broger, O. Mathieu, J. Paszkowski, I. Vaillant, Stress-induced activation of heterochromatic transcription, *PLoS Genet.* 6 (2010) e1001175.
- [110] A. Pecinka, H.Q. Dinh, T. Baubec, M. Rosa, N. Lettner, O. Mittelsten Scheid, Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in *Arabidopsis*, *Plant Cell* 22 (2010) 3118–3129.
- [111] H. Ito, H. Gaubert, E. Bucher, M. Mirozue, I. Vaillant, J. Paszkowski, An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress, *Nature* (2011).



**Figure I-5 : Distribution de la méthylation des gènes chez *A. thaliana* et corrélation avec le taux d'expression**

**A-** Les gènes possèdent principalement de la méthylation CG, sur les exons. Elle est déposée par l'ADN méthyltransférase MET1 et ses cofacteurs VIM1, VIM2 et VIM3.

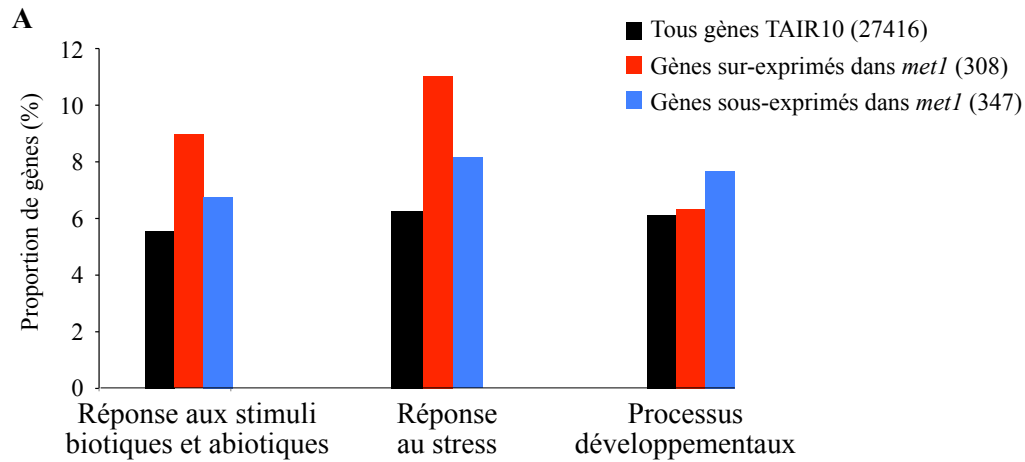
**B-** La distribution de la méthylation CG au niveau des gènes présente une forme parabolique, un taux faible aux extrémités et maximum au centre du corps des gènes. Au contraire, la méthylation CHG et CHH est très faible sur les gènes.

## I.A.5. Régulation épigénétique des gènes

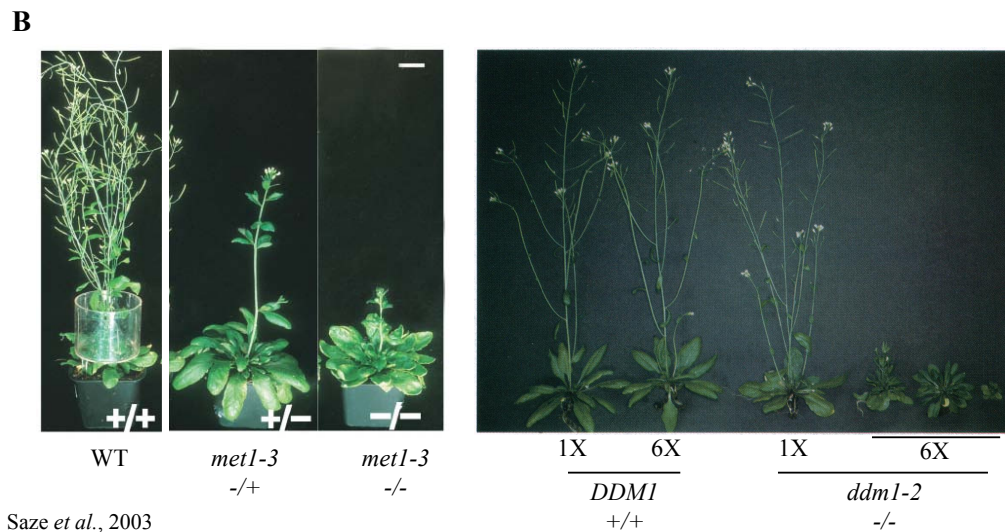
### I.A.4.a. La méthylation ADN

Un tiers des gènes est méthylé, et contrairement aux éléments transposables, cette méthylation est retrouvée presque exclusivement sur les cytosines en contexte CG (**Figure I-5A & B**) (Tran et al. 2005). Elle est déposée sur les exons majoritairement, par l'ADN méthyltransférase METHYLTRANSFERASE 1 (MET1) ainsi que ses cofacteurs VARIANT IN METHYLATION 1 (VIM1), VIM2, et VIM3 (Feng et al. 2010). Sa distribution rappelle celle du positionnement des nucléosomes et suggère que la méthylation ADN puisse avoir un impact dans la définition des exons ou la régulation de l'épissage (Chodavarapu et al. 2010). La protéine de remodelage de la chromatine DECREASE IN DNA METHYLATION 1 (DDM1), bien qu'agissant majoritairement sur les éléments transposables, est aussi nécessaire à la méthylation CG d'environ 50% des gènes euchromatiques méthylés (Zemach et al. 2013). Elle ouvrirait l'accès aux méthyltransférases ADN au niveau des régions très compactées, riches en histone H1.

De façon intéressante, la distribution de la méthylation CGme sur les gènes présente une forme parabolique, elle est déplétée sur les extrémités (2kb du début du gène et 1kb de la fin du gène) et augmente fortement en allant vers le centre du corps du gène (**Figure I-5B**) (Zhang et al. 2006, Zilberman et al. 2007, Cokus et al. 2008). L'expression génique varie de façon opposée à la méthylation ADN sur la région proche du site d'initiation de la transcription (Zilberman et al. 2007, Zemach et al. 2010). Cette absence de méthylation CG aux extrémités serait donc cruciale pour la transcription correcte des gènes. De plus, confirmant cette hypothèse, il a été noté que les gènes ayant une expression modérée sont plus souvent méthylés que les gènes très fortement ou très faiblement transcrits. Les petits gènes méthylés, de moins de 3kb, sont les moins exprimés, leur petite taille leur conférerait donc une plus grande sensibilité à la méthylation ADN. L'analyse de l'expression génique dans le mutant *met1*, déficient en méthylation CG a permis de démontrer également que la méthylation ADN, qu'elle soit présente dans le corps du gène ou dans le promoteur, interfère avec l'élongation de la transcription, et plus le taux de méthylation CG dans la plante sauvage est fort, plus le gène est surexprimé dans le mutant (Zilberman et al. 2007).



D'après les données de Lister *et al.*, 2008



Saze *et al.*, 2003

**Figure I-6 : MET1 et la méthylation CG sont nécessaires à l'expression correcte des gènes, régulant le développement de la plante et la réponse au stress.**

**A- Fraction des gènes dérégulés dans *met1* impliqués dans la réponse au stress et le développement de la plante.** L'ontologie des gènes sur-exprimés et sous-exprimés dans *met1* a été déterminé grâce à l'outil Gene Ontology Annotations de TAIR, en utilisant les données de séquençage ARN à haut-débit réalisé par Lister *et al.*, 2008.

**B- Phénotypes des plantes *met1-3* et *ddm1-2*.** Un retard de floraison est observé dans les plantes hétérozygotes *met1-3* +/- et homozygotes *met1-3* -/- ainsi que dans les générations tardives des mutants *ddm1-2* -/- (1X : autofécondées une fois ; 6X : autofécondées 6 fois).

Dans les mutants déficients en méthylation ADN, de nombreux gènes sont surexprimés ou sous-exprimés (Zhang et al. 2006, Lister et al. 2008, Downen et al. 2012b). Parmi eux, beaucoup sont impliqués dans le développement et la réponse aux stress biotiques et abiotiques, démontrant le rôle fondamental de la méthylation ADN dans ces processus (**Figure I-6A**).

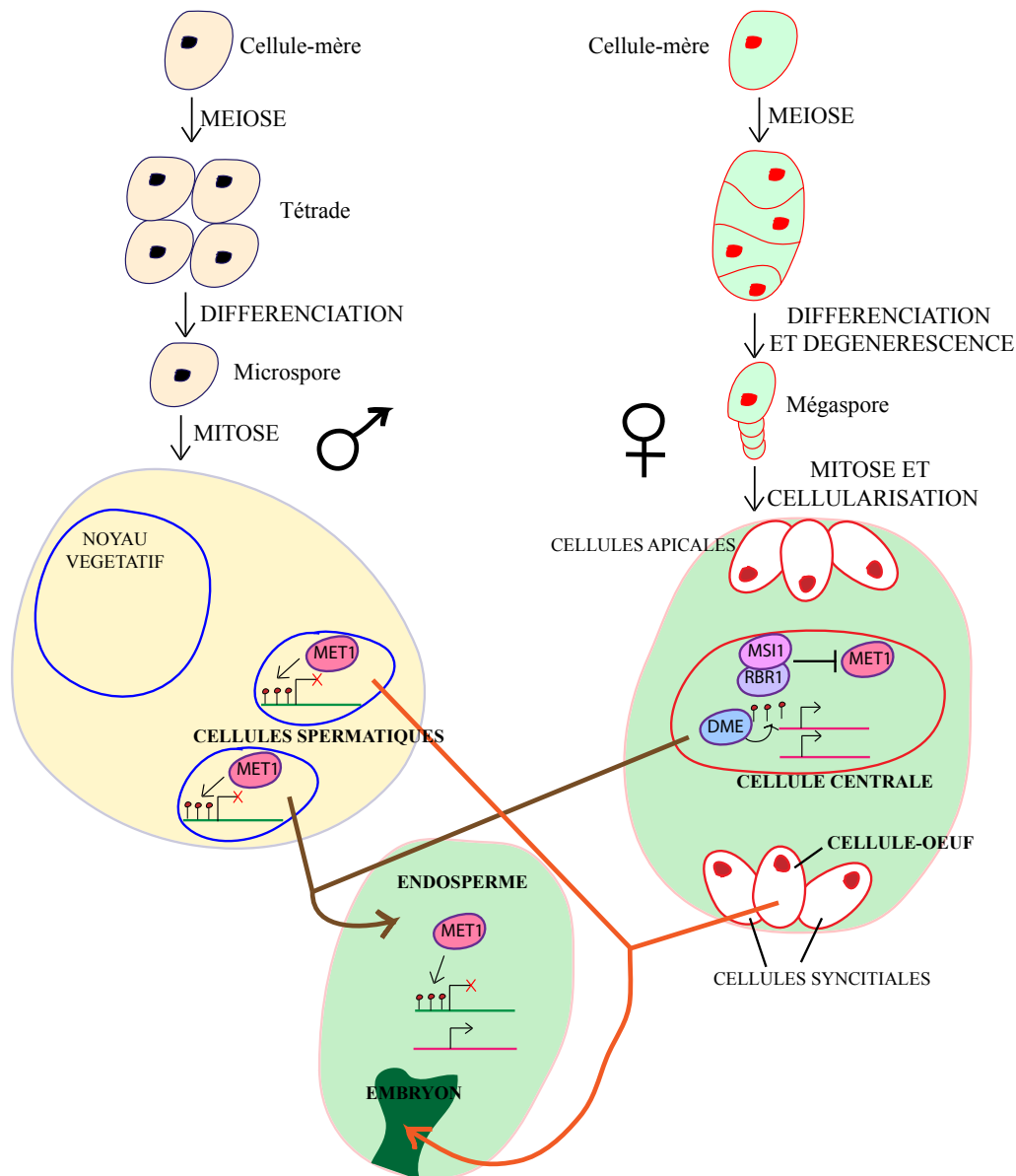
Ceci participe à l'apparition d'anomalies développementales dans ces mutants. Chez le mutant *met1*, où la méthylation CG est perdue, un fort retard de floraison est observé, comme dans les générations tardives du mutant *ddm1* (**Figure I-6B**). La floraison tardive est causée dans ces mutants par l'expression ectopique du gène *FLOWERING WAGENINGEN (FWA)*, due à la déméthylation d'un élément SINE présent dans le promoteur du gène.

L'étude de ce gène, ainsi que des gènes *MEDEA (MEA)*, *FERTILIZATION INDEPENDENT ENDOSPERM (FIE)* et *FERTILIZATION INDEPENDENT SEED 2 (FIS2)* notamment, gènes réprimant le développement de la graine en absence de pollinisation, a permis de mettre en évidence l'implication de la méthylation CG dans le contrôle de l'empreinte parentale (Kinoshita et al. 1999, Vielle-Calzada et al. 1999, Luo et al. 2000, Soppe et al. 2000a, Kinoshita et al. 2004, Jullien et al. 2006, Hsieh et al. 2011). Ce processus permet la régulation de nombreux gènes essentiels au développement normal chez les Mammifères et les Angiospermes. Chez les Plantes, il se produit majoritairement dans l'endosperme dont les fonctions sont analogues au placenta, et affecte 1 à 2% des gènes (Gehring et al. 2011, Hsieh et al. 2011, McKeown et al. 2011, Wolff et al. 2011, Raissig et al. 2013). Quelques gènes présentant une expression monoallélique dans l'embryon dépendante de l'origine parentale ont également été récemment identifiés (Raissig et al. 2013).

### **Comment sont régulés les gènes soumis à empreinte ?**

Le gamétophyte femelle comporte deux cellules, la cellule œuf haploïde et la cellule centrale diploïde (**Figure I-7**). Elles sont fécondées par les deux cellules spermatiques haploïdes du gamétophyte mâle, générant ainsi un embryon diploïde et un endosperme triploïde qui apporte les nutriments nécessaires au développement de l'embryon.





**Figure I-7 : Contrôle de l'expression des gènes soumis à l'empreinte parentale chez *A. thaliana*.**

Chaque cellule-mère subit une méiose puis une différenciation cellulaire, en résulte une cellule haploïde, microspore chez le mâle, et mégaspore chez la femelle. La division de cette cellule est à l'origine de la formation des gamétophytes. La double fécondation de la cellule œuf haploïde et de la cellule centrale diploïde du gamétophyte femelle par les deux cellules spermatisques haploïdes du gamétophyte mâle génère un embryon diploïde et un endosperme triploïde.

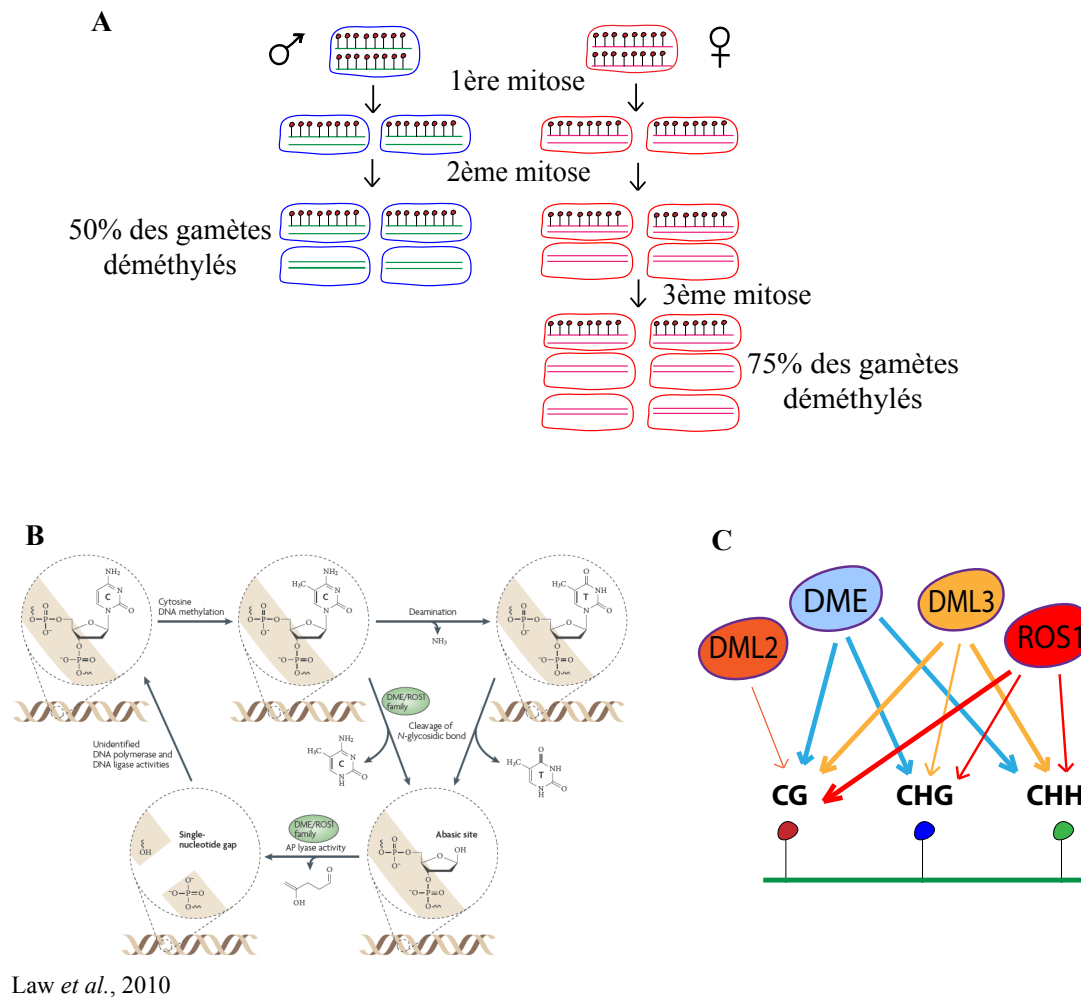
Dans le gamétophyte mâle, la présence de la méthyltransférase ADN MET1 entraîne la méthylation des promoteurs de gènes soumis à l'empreinte parentale et ainsi leur répression. Ces gènes sont exprimés dans le gamétophyte femelle, grâce à la présence d'un complexe incluant les protéines RBR1 et MSI1 qui réprime la transcription du gène MET1, et de la glycosylase ADN DME participant à leur déméthylation active. Après fécondation, l'expression de cet allèle maternel est maintenue dans l'endosperme, tandis que l'allèle paternel demeure méthylé par MET1 et réprimé.

Dans le gamétophyte femelle, la transcription du gène *MET1* est réprimée par le complexe formé par la protéine impliquée dans la prévention de la prolifération syncytiale incontrôlée, RETINOBLASTOMA RELATED 1 (RBR1) et l'homologue de RbAp48, MULTICOPYSUPPRESSOR OF IRA1 (MSI1), qui empêche la production d'endosperme à partir de cellules centrales non fécondées (Jullien et al. 2008). La répression de *MET1* ainsi que l'action de la déméthylase ADN DEMETER (DME) dans la cellule centrale permet la transcription de gènes tels que *MEDEA*, *FWA* ou *FIS2* (**Figure I-7**) (Choi et al. 2002, Gehring et al. 2006, Jullien et al. 2006, Jullien et al. 2008). L'expression de cet allèle maternel est maintenue dans l'endosperme, après fécondation, tandis que l'allèle paternel est réprimé par *MET1* qui dépose de la méthylation CG sur le promoteur de ces gènes, comme dans le gamétophyte mâle et durant la phase végétative.

L'empreinte parentale dans l'embryon, elle, est moins bien documentée à ce jour mais nécessiterait la suppression et la remise en place des marques épigénétiques entre les générations.

Dans l'endosperme, ce processus pourrait provenir du conflit d'intérêt parental pour l'allocation des ressources, comme chez les Mammifères. Les gènes ciblés dans l'embryon pourraient, eux, participer à l'interface mère-nourrisson, ce qui est concordant avec le fait que la très grande majorité des gènes identifiés sont exprimés maternellement. Mais ceci demeure hypothétique en l'état actuel des connaissances (Raissig et al. 2013).

Le rôle crucial de la méthylation CG dans le contrôle de l'expression des gènes durant le développement embryonnaire se révèle également par le phénotype des graines issues du croisement de mutants *met1* avec des plantes sauvages (FitzGerald et al. 2008). En effet, lorsque *met1* est transmis par le gamétophyte mâle, des graines de taille réduite sont obtenues, ceci est probablement lié à l'absence de répression de gènes soumis à empreinte codant des inhibiteurs de croissance. La transmission de la mutation *met1* dans le gamétophyte femelle entraîne au contraire la formation de grosses graines. Cela serait dû à une prolifération excessive, non pas du gamétophyte ou de l'embryon lui-même, mais des tissus sporophytiques maternels entourant le sac embryonnaire. Ainsi, *MET1* contrôle la croissance de l'endosperme mais réprime également les gènes maternels impliqués dans la division cellulaire et l'élongation des intéguments, limitant ainsi la croissance de la graine (FitzGerald et al. 2008).



**Figure I-8 : Processus de déméthylation passive et active chez *A.thaliana***

**A- Modèle de déméthylation passive.** En absence de maintenance de la méthylation ADN durant la gaméto-genèse, tel que dans le mutant *met1*, l'ADN est passivement déméthylé au cours des deux ou trois mitoses. En résulte l'obtention de gamètes hémiméthylés ou totalement déméthylés.

**B- Mécanisme de déméthylation active.** Chez Arabidopsis, l'ADN est déméthylé activement via l'action de glycosylases bifonctionnelles de la famille de DME. Tout d'abord, la cytosine méthylée est libérée par clivage de la liaison N-glycosidique, générant un site abasique. Puis la liaison phosphodiester est supprimée au niveau des extrémités 3' et 5' du site abasique via l'activité de lyase apyrimidique, ceci génère une brèche d'un nucléotide, réparée par une polymérase et une ligase ADN.

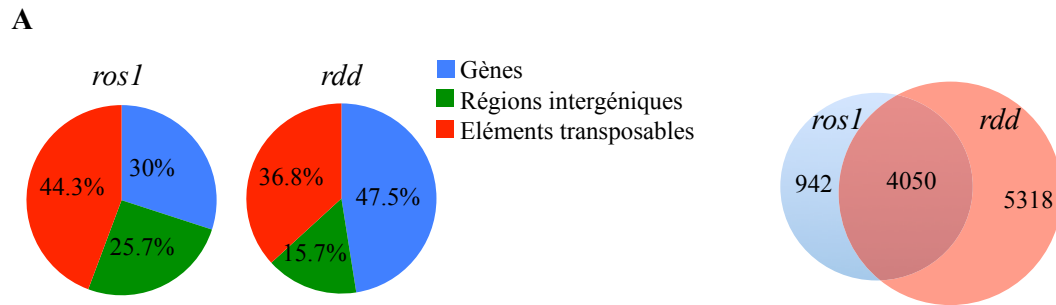
**C- Déméthylases ADN décrites chez *A. thaliana*.** Chez Arabidopsis, quatre déméthylases ADN ont été décrites, incluant DME, ROS1 et DML3, capables de déméthyliser les cytosines dans les trois contextes. Seule une faible activité sur les sites CG a par contre été détectée pour DML2. La largeur des flèches corrèle avec la préférence des déméthylases ADN pour chaque contexte.

### ***I.A.4.b. La déméthylation ADN***

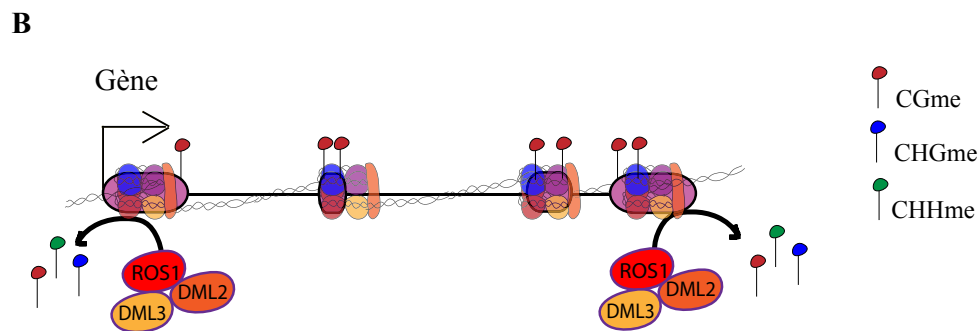
La méthylation ADN peut être effacée de façon passive, au cours des réplifications, par l'absence de maintenance de la méthylation mais aussi de façon active, chez les Plantes, par l'action de déméthylases ADN (**Figure I-8A**). Ces protéines sont à la fois des glycosylases ADN et des lyases ADN de la superfamille HhH-GPD (Helix-hairpin-Helix). Leur activité glycosylase permet le clivage de la liaison N-glycosydique de la 5-méthylcytosine (5-meC) par  $\beta,\delta$ -élimination générant un site abasique au niveau duquel l'ADN est ensuite coupé par l'activité lyase (**Figure I-8B**) (Agius et al. 2006). Une AP endonucléase crée une extrémité 3'OH permettant à une ADN polymérase de réparation d'insérer un nucleotide non méthylé (Kapoor et al. 2005). Enfin, une ligase ADN rebouche la brèche. Récemment, Martinez-Macias *et al.* ont identifié une ADN 3' phosphatase, ZINC FINGER DNA 3'-PHOSPHATASE (ZDP), opérant sur les produits d'incision de REPRESSOR OF SILENCING 1/DEMETER-LIKE 1 (ROS1/DML1) (Martinez-Macias et al. 2012). Ils ont également démontré que la protéine X-RAY REPAIR CROSS COMPLEMENTING (XRCC1) interagit avec ROS1 et ZDP et stimule leur activité enzymatique, elle facilite également la ligation au niveau de la brèche (Martinez-Macias et al. 2013).

Chez *Arabidopsis*, quatre ADN déméthylases ont été décrites, la protéine DME ainsi que ses paralogues ROS1, DEMETER-LIKE 2 (DML2) et 3 (DML3) (**Figure I-8C**) (Gehring et al. 2006, Morales-Ruiz et al. 2006, Penterman et al. 2007b, Ortega-Galisteo et al. 2008). Ces ADN déméthylases sont exprimées dans les organes principaux de la plante, racines, tiges, inflorescences ainsi que feuilles caulines. Des expériences *in vitro* ont démontré que DME est capable de déméthyliser les 5-meC quelque soit leur contexte, ROS1 et DML3 également même s'ils présentent une affinité plus forte respectivement pour les sites CG, et CG et CHH. DML2, quant à elle, a une activité *in vitro* beaucoup plus faible et détectée seulement en présence d'oligonucléotides méthylés en contexte CG (**Figure I-8C**).

L'étude des simples mutants *ros1*, *dml2* et *dml3* ainsi que du triple mutant *ros1 dml2 dml3* (*rdd*) a démontré qu'ils agissent de façon redondante sur certains loci mais ont également chacun des cibles spécifiques (**Figure I-9A**) (Penterman et al. 2007b, Lister et al. 2008).



Adapté de Qian *et al.*, 2012



**Figure I-9 : Caractéristiques des régions cibles des ADN déméthylases.**

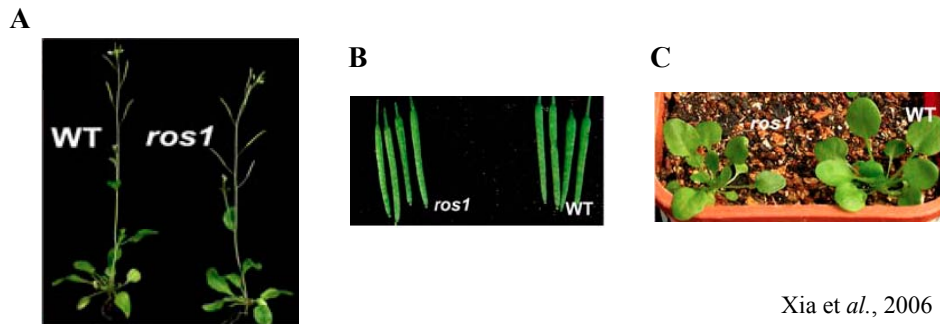
**A-** Distribution des régions différenciellement méthylées dans le mutant *ros1* et le triple mutant *ros1 dml2 dml3* (*rdd*), et chevauchement de ces régions entre le triple et le simple mutant.

**B-** Les déméthylases ADN ROS1, DML2 et DML3 ciblent les extrémités 3' et 5' des gènes afin de prévenir l'accumulation de méthylation ADN qui pourrait affecter leur transcription.

De plus, l'analyse récente des profils de méthylation ADN sur le génome entier a mise en évidence 5000 loci hyperméthylés dans le mutant *ros1-4* et 9000 dans le triple mutant *rdd*, incluant gènes, transposons et zones intergéniques (Qian et al. 2012). 4000 de ces loci sont altérés dans *ros1* et *rdd*, ce qui suggère un rôle prépondérant de ROS1 sur 40% des loci cibles des déméthylases (**Figure I-9A**).

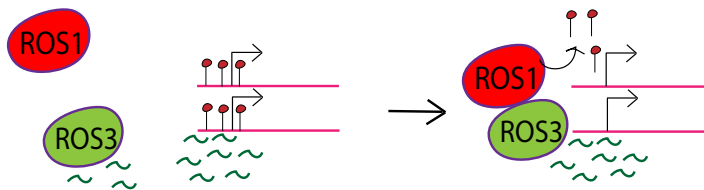
De façon intéressante, alors que les éléments transposables sont le plus souvent hyperméthylés, dans le mutant *rdd*, sur la totalité de leur séquence, l'hyperméthylation au niveau des gènes est retrouvée principalement sur les extrémités 3' et 5'. Les déméthylases seraient donc responsables de la déplétion en méthylation ADN observée sur ces zones dans les plantes sauvages (Penterman et al. 2007b, Lister et al. 2008) (**Figures I-5B et I-9B**). De plus, dans ce triple mutant, la zone hyperméthylée tend à être enrichie en petits ARN de 24 nt, ce qui suggère que l'hyperméthylation fait intervenir la voie RdDM, voie de silencing par les petits ARN (Penterman et al. 2007a, Qian et al. 2012). De façon concordante, l'hyperméthylation des loci dans le mutant *ros1* implique des acteurs de cette voie de silencing tels que Pol IV, RNA-DEPENDENT RNA POLYMERASE 2 (RDR2), DICER-LIKE 3 (DCL3) et ARGONAUTE 4 (AGO4) (Penterman et al. 2007a). Dans les mutants RdDM, le gène *ROS1* est fortement réprimé et la protéine DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1), autre composant de cette voie, est nécessaire non seulement à la méthylation de certains loci sur les trois contextes mais également à la déméthylation CG de ceux-ci (Kanno et al. 2005, Huettel et al. 2006). Ainsi, ROS1, DML2 et DML3 contrecarreraient l'action de la voie RdDM afin qu'elle n'empiète pas sur les gènes, voie RdDM qui elle-même régule leur fonctionnement.

Exceptés certains transgènes, gènes et éléments transposables dont le silencing est contrôlé par l'action de ROS1, l'expression des loci ciblés n'est, dans la grande majorité des cas, pas affectée par l'hyperméthylation apparaissant dans les mutants. Ainsi, le séquençage ARN à haut-débit dans les bourgeons floraux indique que 0,53% (146 gènes) des gènes sont affectés transcriptionnellement dans *rdd*, seuls 10 d'entre eux présentent une hyperméthylation significative et sont donc probablement des cibles des déméthylases ADN. Une forte stringence dans l'analyse de la méthylation ADN dans cette étude ou un effet indirect des déméthylases ADN pourrait expliquer la différence d'expression observée sur les autres loci.



**Figure I-10 : Phénotypes des plantes *ros1*.**

Les plantes entières (A), les siliques (B) ainsi que les plantules (C) sauvages (WT) et mutantes pour le gène *ROS1* sont représentées.



**Figure I-11 : Mécanisme putatif de guidage de la déméthylase ADN ROS1 par la protéine ROS3.**

La protéine ROS3 se lierait aux petits ARN et permettrait le recrutement de ROS1 au niveau des loci homologues, et ainsi leur déméthylation.

De plus, moitié d'entre eux sont plus transcrits dans le triple mutant qu'en fond sauvage, contrairement à l'hypothèse conférant aux déméthylases un rôle anti-silencing (Penterman et al. 2007b, Zhu et al. 2007, Lister et al. 2008 ).

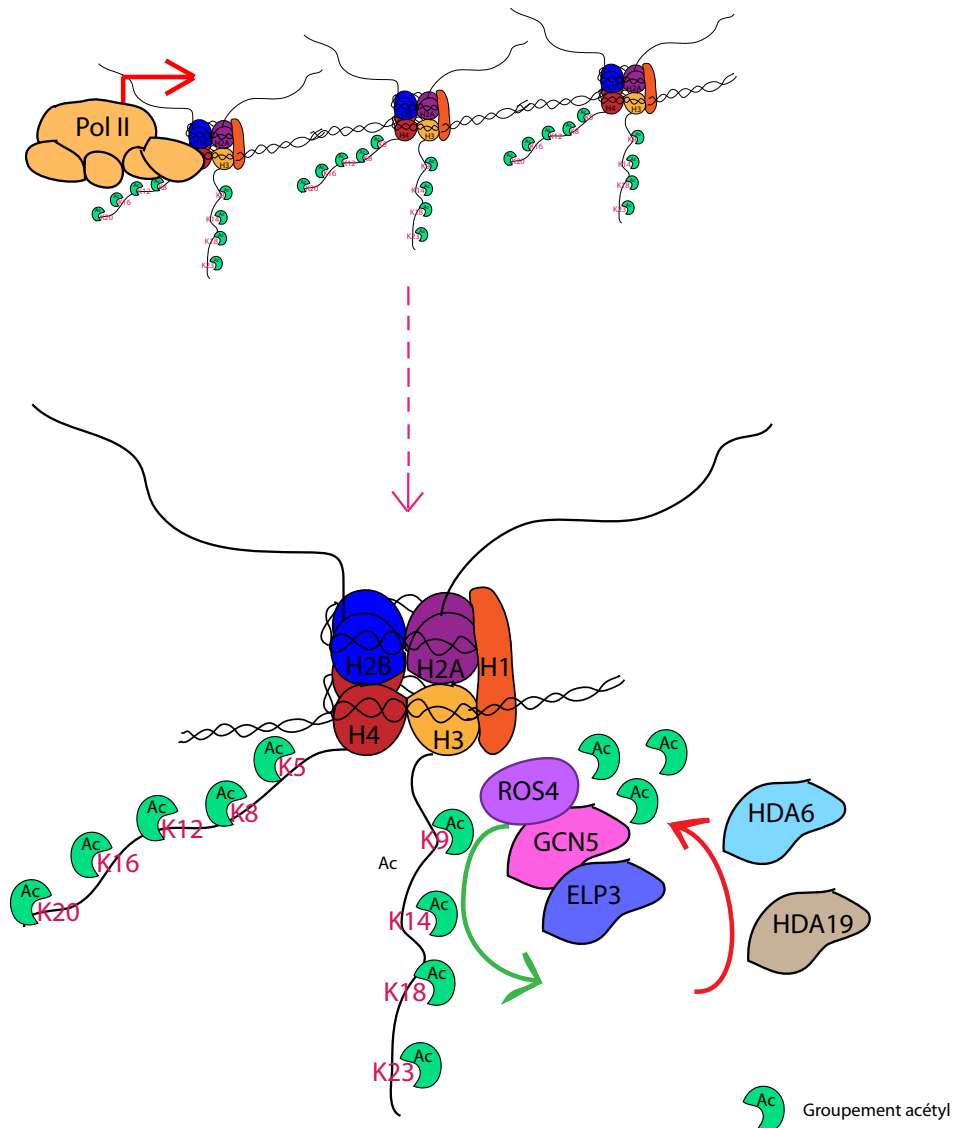
De même que l'effet limité sur l'expression des gènes cibles, l'absence d'anomalies phénotypiques des simples et du triple mutant, du moins dans les premières générations, suggère que ces protéines ne sont pas essentielles au développement correct de la plante (**Figure I-10**) (Gong et al. 2002, Penterman et al. 2007b). Cependant, un article récent a démontré que ROS1 contrôlerait la taille de la population de cellules souches stomatales, par la déméthylation du promoteur du gène EPF2, le mécanisme de déméthylation ADN active est donc responsable de certaines étapes clés du développement (Yamamuro et al. 2014). ROS1 participerait aussi à la réponse immune induite par la plante lors d'infection bactérienne, en amorçant l'activation transcriptionnelle de certains gènes de défense et de TEs (Yu et al. 2013). Ainsi ces déméthylases pourraient participer à maintenir le génome plastique pour une réponse rapide au stress.

### **Comment et pourquoi certains loci sont-ils spécifiquement ciblés par les déméthylases ADN ?**

Ceci demeure très flou. Les différentes études réalisées ont établi que ces loci sont distribués de façon homogène sur les cinq chromosomes, sont enrichis en petits ARN de 24 nt, et chevauchent pour la plupart des séquences répétées (Penterman et al. 2007a, Qian et al. 2012 ). L'identification de la protéine REPRESSOR OF SILENCING 3 (ROS3), colocalisant avec ROS1 et se liant à des petits ARN *in vitro* apporte une piste, cette protéine pourrait guider ROS1 vers ses cibles grâce à sa liaison à des petits ARN homologues (**Figure I-11**) (Zheng et al. 2008).

La protéine DME est, elle, impliquée dans le phénomène d'empreinte parentale, en association avec la méthyltransférase ADN MET1. L'étude des profils de méthylation ADN dans l'embryon et l'endosperme ainsi que d'endosperme de plantes mutées pour *DME* a également révélé que cette déméthylase ADN est responsable de la réduction en méthylation ADN principalement au niveau de sites CG dans l'endosperme. Au niveau de certains gènes, la perte en méthylation CG est localisée à l'extrémité 5' et provoque leur activation transcriptionnelle. Ces analyses ont donc permis de révéler de potentiels nouveaux gènes soumis à empreinte (Gehring et al. 2011, Hsieh et al. 2011, McKeown et al. 2011, Wolff et al. 2011, Raissig et al. 2013).





**Figure I-12 : Acétylation des histones H3 et H4.**

L'acétylation des histones est associée à la transcription. Elle affecte principalement les lysines 9, 14, 18 et 23 de l'histone H3 ainsi que les lysines 5, 8, 12, 16 et 20 de l'histone H4. ROS4, GCN5 et ELP3 contribueraient au dépôt des groupements acétyl tandis que des déacétylases telles que HDA6 et HDA19 sont capables d'enlever ces groupements.

Parallèlement à la méthylation et la déméthylation ADN, de nombreuses modifications post-traductionnelles des histones ainsi que l'incorporation de variants d'histones ont un rôle crucial dans la compaction de la chromatine ainsi que la modulation des profils d'expression.

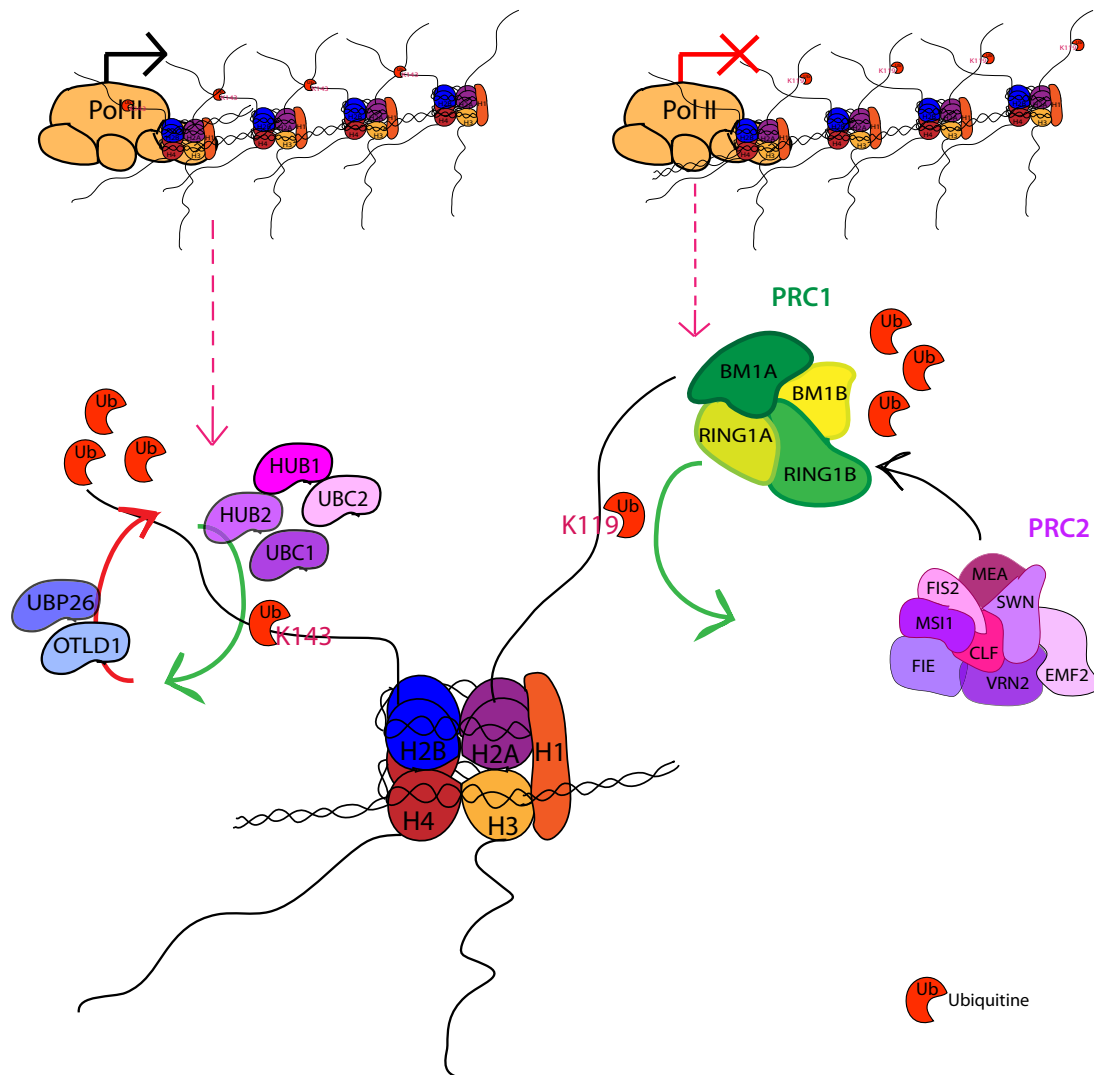
#### ***I.A.4.c. Les marques histones et variants d'histones***

##### **I.A.4.a.1. L'acétylation et la déacétylation des histones**

L'acétylation des histones consiste en l'ajout d'un groupement acétyl sur un résidu Lysine de l'histone par une histone acétyltransférase (HAT), elle neutralise la charge positive de la lysine affectée et ainsi peut modifier les interactions de l'histone avec la molécule d'ADN ou avec d'autres protéines (**Figure I-12**). Chez *Arabidopsis*, l'acétylation peut affecter les lysines 9, 14, 18, 23 et 27 de l'histone H3 ainsi que les lysines 5, 8, 12, 16 et 20 de l'histone H4. Cette modification d'histones est associée à une activation de la transcription, aux processus de réplication, recombinaison ou réparation. Les HAT sont classifiées en quatre familles, GCN5-RELATED N-ACETYLYTRANSFERASE (GNAT), MYST, CREB BINDING PROTEIN (CBP) /p300 et TBP-ASSOCIATED FACTOR 1 (TAF1)/TAFII250. La famille GNAT inclut notamment la protéine GCN5 / HISTONE ACETYLTRANSFERASE OF THE GNAT FAMILY 1 (HAG1), HAT majeure chez *Arabidopsis* nécessaire à la transcription d'un grand nombre de gènes, et la protéine ELONGATOR PROTEIN 3 (ELP3) /ELO3/HAG3 qui favorise l'élongation de la transcription par interaction avec l'ARN polymérase. La mutation des gènes codant pour ces HAT provoque l'apparition de nombreux défauts dans le développement et la croissance de la plante (Bertrand et al. 2003, Bertrand et al. 2005, Berr et al. 2011).

Afin de maintenir un équilibre constant dans le niveau d'acétylation des histones, des histones déacétylases (HDAC) participent à la suppression des groupements acétyl (Berr et al. 2011). Il existe quatre sous-classes, RDP3, HDA1 et SIR2, homologues à la Levure, et HD2, spécifique des Plantes (Pandey et al. 2002) (**Figure I-12**). La mutation du gène codant la protéine HISTONE DEACETYLASE 19 (HDA19) /HD1, appartenant à la classe RDP3 entraîne des anomalies pléiotropiques, corrélée à l'expression ectopique de certains gènes tels que *SUPERMAN* (*SUP*) (Wu et al. 2000, Tian et al. 2001, Tian et al. 2003).

7% des gènes sont affectés transcriptionnellement dans ce mutant, une partie de ces gènes sont surexprimés et déacétylés au niveau de leur promoteur, d'autres sont sous-exprimés. Ceci suggère un rôle des HDAC dans l'expression correcte de nombreux gènes, qu'elles agissent de façon directe ou indirecte sur les gènes cibles (Tian et al. 2005).



**Figure I-13 : Ubiquitination des histones H2A et H2B.**

L'ubiquitination affecte la lysine 143 de l'histone H2B et la lysine 119 de l'histone H2A. Les protéines HUB1 et 2 et UBC1 et 2 coopèrent pour le dépôt d'ubiquitine sur l'H2B tandis que UBP26 et OTLD1 enlèvent le groupement ubiquitine sur des loci hétérochromatiques notamment. La marque H2B<sup>Ub</sup> est associée à des gènes transcrits. Au contraire, l'H2A<sup>Ub</sup> tend à inhiber la transcription. Elle est déposée par le complexe PRC1. Celui-ci recruterait ensuite le complexe PRC2 qui maintiendrait l'état réprimé par le dépôt d'H3K27me3.

Parmi la classe RDP3, HDA6 a également une fonction importante dans le silencing des éléments transposables et des gènes ARNr ainsi que dans la maintenance de la méthylation ADN (Lippman et al. 2003a, Earley et al. 2006, Earley et al. 2010). Son action implique la voie RdDM et la méthyltransférase MET1 avec laquelle elle est capable d'interagir. HDA6 régule également la méthylation ADN au niveau de certains promoteurs, via un mécanisme indépendant (O'Neill et al. 1998, Stroud et al. 2013). Elle pourrait déacétyler H3K14, H4K5 et H4K12 (Earley et al. 2006).

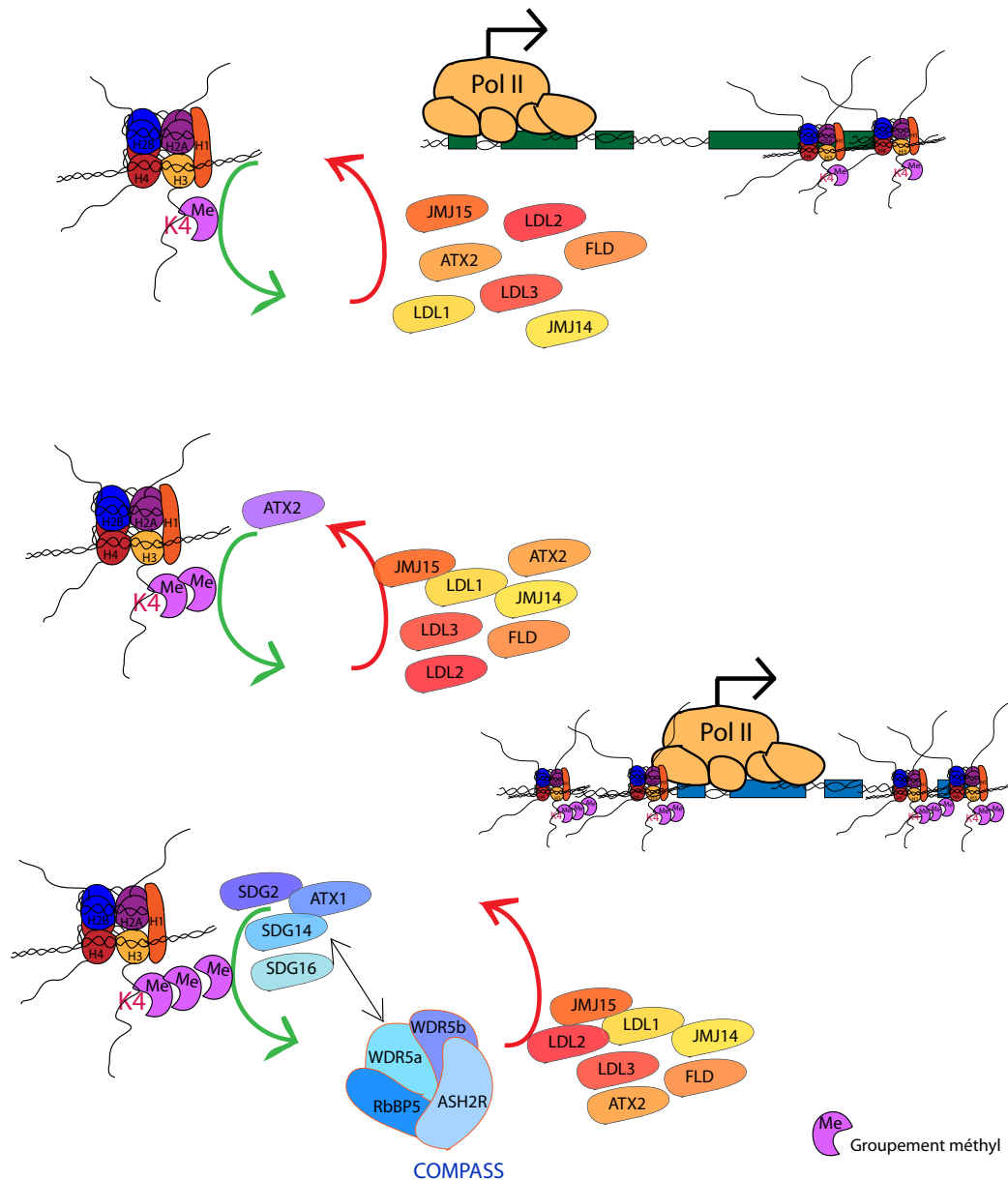
#### **I.A.4.a.2. L'ubiquitination et déubiquitination des histones**

L'ubiquitination des histones se caractérise par la liaison d'une ubiquitine. Cette petite protéine est tout d'abord activée et liée à l'enzyme E1, puis conjuguée à l'enzyme E2, qui se lie elle-même à l'enzyme E3 permettant le transfert de l'ubiquitine sur un résidu lysine (Pickart 2001). Plusieurs histones et variants d'histones peuvent être monoubiquitinés, cependant les études se sont jusqu'alors majoritairement focalisées sur les histones H2A et H2B (**Figure I-13**). Les protéines BMI1A, BMI1B, RING1A et RING1B appartenant au complexe Polycomb Repressive Complex 1 (PRC1) participent à la monoubiquitination de la lysine 119 de l'histone H2A (Bratzel et al. 2010). Cette marque bloque l'initiation de la transcription des loci ciblés et a un rôle important dans la régulation de l'activité du méristème apical, la germination des graines, la répression des gènes de régulation embryonnaire durant la phase végétative. De façon intéressante, alors que chez les Animaux le dépôt d'H3K27me3 par POLYCOMB REPRESSIVE COMPLEX 2 (PRC2) précède le recrutement de PRC1, une étude a démontré que, sur les gènes analysés, codant pour des protéines de maturation des graines, le dépôt d'H2AUb initie la répression transcriptionnelle, ensuite maintenue par la triméthylation de l'H3K27 (Yang et al. 2013).

La monoubiquitination de l'histone H2B (H2BUb) est, elle, retrouvée de la Levure aux Animaux, sur une lysine très conservée, la lysine 143 chez Arabidopsis (**Figure I-13**). Cette marque est associée à des gènes actifs et se localise souvent sur des régions également enrichies en H3K4me3 et/ou H3K36me3 (Roudier et al. 2011).

Quatre protéines participeraient au dépôt de cette marque chez Arabidopsis, UBIQUITIN CARRIER PROTEIN 1 (UBC1) et UBC2 agiraient de façon redondante, au côté de HISTONE MONOUBIQUITINATION 1 (HUB1) et HUB2.

L'H2BUb est impliquée dans la régulation de la floraison, du rythme circadien, du cycle cellulaire, de la photomorphogénèse et dans la défense de la plante contre les pathogènes.



**Figure I-14 : Méthylation de l'H3K4.**

La lysine 4 de l'histone H3 peut être mono-, di-, ou triméthylée. L'H3K4me1 est enrichie sur les gènes de grande taille, principalement en 3', tandis que l'H3K4me2 et me3 sont retrouvées à l'extrémité 3' de gènes mais aussi dans les promoteurs. ATX2 pourrait diméthyliser l'H3K4. SDG2, SDG14, SDG16 et AXT1 pourraient participer à la triméthylation de cette lysine via leur interaction avec un sous-complexe COMPASS. Un grand nombre de déméthylases histone seraient capables de cibler l'H3K4.

Plusieurs études ont démontré l'interaction génétique ou physique d'ubiquitinasés avec des gènes ou protéines impliquées dans la transcription et le recrutement de la polymérase ARN II telles que le complexe Polymerase Associated Factor 1 (PAF1), le complexe Mediator ou la chaperone d'histones FACILITATES CHROMATIN TRANSCRIPTION (FACT) qui déstabilise les nucléosomes pour faciliter le passage de la Pol II (Dhawan et al. 2009, Gu et al. 2009, Lolas et al. 2010). Cependant, l'effet exact d'H2Bub dans la transcription reste flou.

Les taux d'H2BUB sont régulés par l'action d'enzymes de débubiquitination, la protéase ubiquitine UBIQUITIN-SPECIFIC PROTEASE 26 (UBP26)/SUP32 impliquée dans la répression de loci hétérochromatiques ainsi que dans la régulation de la floraison, et la débubiquitinase OTUBAIN-LIKE DEUBIQUITINASE 1 (OTLD1) identifiée récemment (Sridhar et al. 2007, Schmitz et al. 2009) (**Figure I-13**).

#### **I.A.4.a.3. La méthylation/déméthylation des histones**

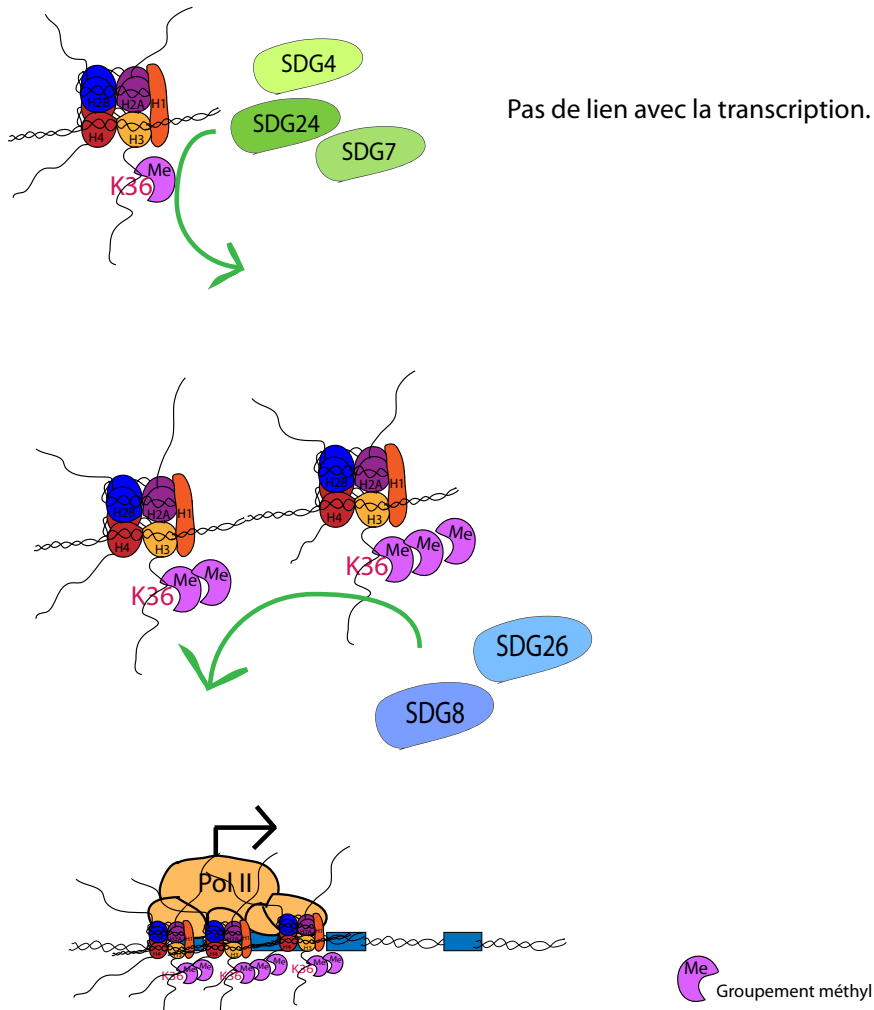
Cette modification de queues d'histones consiste en l'ajout d'un, deux ou trois groupement(s) méthyl sur un résidu lysine ou arginine par une histone méthyltransférase (HMT). Chez Arabidopsis, toutes les HMT identifiées possèdent un domaine catalytique SET (du nom de l'HMT identifié chez la Drosophile, Suppressor of variegation (Su(var)3-9), Enhancer of Zeste (E(z)) and Trithorax (TRX)). Elles sont au nombre de 47 mais seules quelques unes d'entre elles ont été caractérisées.

Les gènes sont majoritairement enrichis en mono-, di- et triméthylation de la Lysine 4 de l'histone H3 (H3K4me1/me2/me3), en H3K36me1/2/3, en H3K27me1/2/3 et en H3K9me3.

#### **La méthylation de l'H3K4**

La méthylation de la lysine 4 de l'histone H3 touche 80% à 90% des gènes. L'H3K4me1 est retrouvée dans les régions transcrites avec un biais en 3', comme la méthylation CG, et est enrichie dans les gènes plus longs (**Figure I-14**). H3K4me2 et H3K4me3 se localisent dans les promoteurs et l'extrémité 3' des gènes. La triméthylation de la lysine 4 marque des gènes fortement et largement exprimés, tandis que l'H3K4me2 colocalise en partie avec l'H3K27me3 et s'associerait préférentiellement à des gènes ayant une expression tissu-spécifique (Zhang et al. 2009, Roudier et al. 2011).

Des analyses phylogénétiques et bioinformatiques ont identifiés dix H3K4 histone méthyltransférases putatives. Parmi elles, les protéines SDG2/ATXR3 et ATX1/SDG27 sont capables de trimétyler l'H3K4 (**Figure I-14**). Le complexe fortement conservé COMPASS est également impliqué dans le dépôt d'H3K4me3.



**Figure I-15 : Méthylation de l'H3K36.**

Contrairement à la monométhylation d'H3K36, les marques H3K36me2 et me3 corrélient avec la transcription. SDG4, 24 et 27 seraient capables de déposer l'H3K36me1, SDG8 et SDG26 l'H3K36me2 et me3.

Il comprend, chez *Arabidopsis*, les protéines ARABIDOPSIS ASH2 RELATIVE (ASH2R), RETINOBLASTOMA BINDING PROTEIN 5 (RbBP5) et WD40 REPEAT 5a (WDR5a) et WDR5b.

Il a été démontré que ASH2R, RbBP5 et WDR5a forment un sous-complexe servant de plateforme à la méthylation H3K4me3, comme observé précédemment chez les Mammifères, via l'interaction de WDR5a avec les HMT ATX1, SET DOMAIN PROTEIN 14 (SDG14) et SDG16 (Jiang et al. 2009, Jiang et al. 2011). Dans les mutants *sdg2*, *atx1* et *ash2r*, de nombreux gènes sont dérégulés et les plantes présentent des anomalies phénotypiques, cette marque est donc nécessaire au développement correct de la plante.

A ce jour, seule la protéine ATX2/SDG30 a été identifiée comme une potentielle histone méthyltransférase H3K4me2 (**Figure I-14**). La mutation du gène *ATX2* n'a pas d'impact sur le phénotype des plantes, d'autres protéines pourraient avoir une fonction redondante à ATX2.

Enfin, aucune HMT capable de déposer l'H3K4me1 n'a été mise en évidence chez *Arabidopsis*.

De façon intéressante, l'HMT SDG8/EFS/CCR1/ASSH2 serait capable de se lier à l'H3K4me grâce à son domaine Cystéine/Tryptophane (CW) et de méthyler cette même histone au niveau de la lysine 36, preuve d'une coopération des marques histones dans la régulation du génome (Hoppmann et al. 2011).

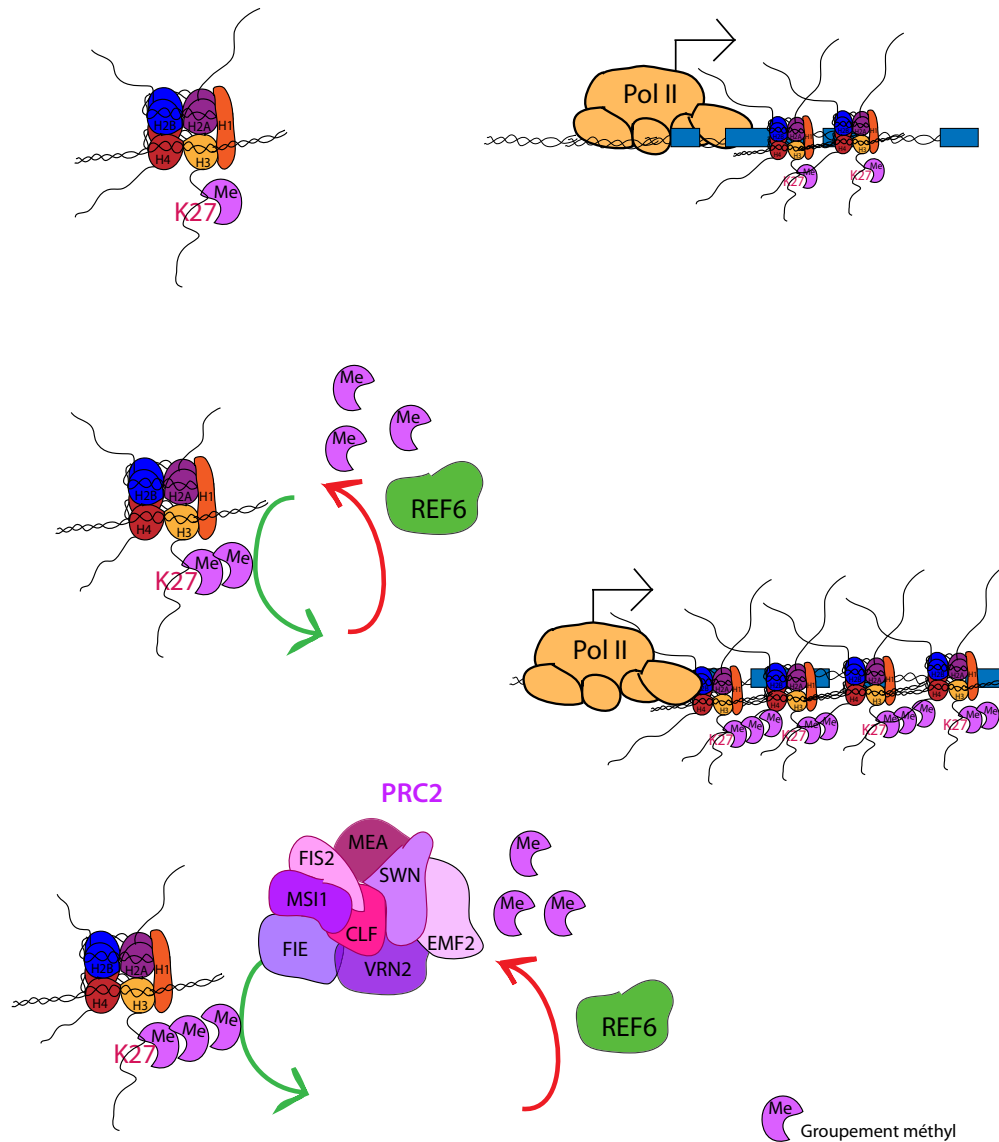
### La méthylation de l'H3K36

L'H3K36me2 et l'H3K36me3 sont notamment retrouvées sur les gènes MADS-box impliqués dans le contrôle de la floraison et du développement floral (**Figure I-15**). Elles corréleront avec la transcription et sont localisées dans la région transcrite des gènes et enrichie au niveau de l'extrémité 5'. Comme l'H3K4me3, ces marques sont associées à des gènes fortement exprimés mais également à des gènes ayant une expression tissu-spécifique (Xu et al. 2008, Grini et al. 2009).

Comme SDG8, SDG26 est impliquée dans le dépôt de groupements méthyl sur l'H3K36, mais ces protéines régulent de façon indépendante l'expression des gènes (Xu et al. 2008).

La baisse en H3K36me2/3 dans le mutant *sdg8* s'accompagne d'une hausse en H3K36me1, marque non associée à la transcription. L'H3K36me2 et me3 pourraient agir en lien avec l'H3K27me3 pour réprimer certains gènes, comme le suggère la mise en évidence d'une interaction épigénétique entre le mutant *sdg8* et le mutant *curly leaf (clf)* du complexe PRC2, déficient en H3K27me3. Selon des analyses de conformation de protéines, SDG4, SDG7 et SDG24 pourraient participer à la monométhylation d'H3K36 (Xu et al. 2008) (**Figure I-15**).





**Figure I-16 : Méthylation de l'H3K27.**

La monométhylation d'H3K27 est retrouvée au centre des gènes ciblés alors que di- et triméthylation se distribuent sur toute la longueur des gènes. Le complexe PRC2 est impliqué dans le dépôt d'H3K27me3 tandis que la déméthylase histone REF6 pourrait cibler les marques H3K27me2 et me3.

### La méthylation de l'H3K27

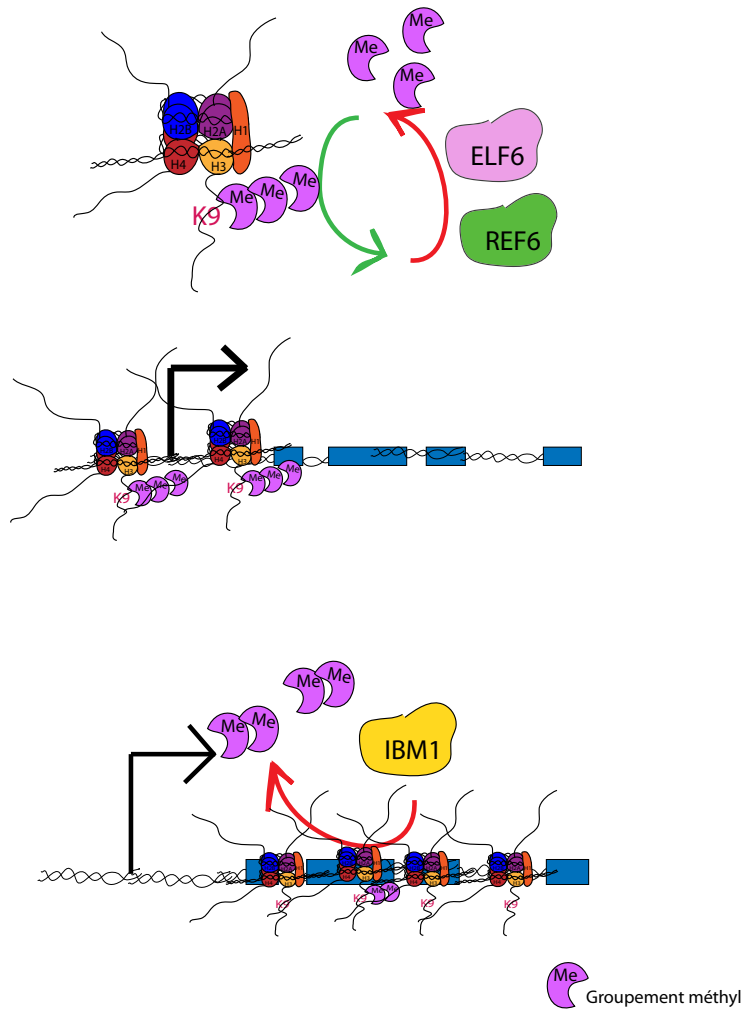
La marque H3K27me3 est très présente au niveau de nombreux gènes (**Figure I-16**). Elle est antagoniste à la marque H3K4me3 active. Ces deux marques pourraient s'associer sur certains loci et former des domaines bivalents, comme chez les Mammifères, pour permettre une activation rapide de certains gènes régulateurs clés, leur association sur la même fibre chromatinienne a été observée récemment sur une sous-population de cellules dans de jeunes plantules (Jiang et al. 2008 , Berr et al. 2010 , Roudier et al. 2011 , Sequeira-Mendes et al. 2014).

Le complexe PRC2, déposant cette marque, inclut les protéines FIE, et MSI1 mais également d'autres cofacteurs dont la présence est variable, EMBRYONIC FLOWER2 (EMF2), FIS2, VERNALIZATION2 (VRN2), MEA, CLF et SWINGER (SWN) (**Figure I-16**). Ainsi, trois complexes PRC2 pourraient se former chez Arabidopsis et auraient des rôles distincts, dans la vernalisation, dans le développement correct de l'endosperme et l'embryon, dans le développement végétatif mais également dans le phénomène d'empreinte parentale dans l'endosperme et dans l'embryon. CLF, SWN et MEA porteraient l'activité d'histone méthyltransférase H3K27 dans l'endosperme (He et al. 2013). Cette marque est très dynamique, déposée puis enlevée tout au long des étapes du développement. La présence de Polycomb Response Elements (PREs), identifiés au niveau de certains loci cibles, pourrait permettre le recrutement du complexe PRC2 (Schwartz et al. 2008). Des ARN non-codants pourraient également guider le complexe, tels que l'ARN COLDAIR sur le gène *FLOWERING LOCUS C (FLC)* interagissant avec les protéines CLF et SWN (Jiang et al. 2008).

La mono- et la diméthylation d'H3K27 sont, elles, retrouvées sur de nombreux éléments transposables mais aussi sur 30% des gènes.

Tandis que H3K27me1 est plus fortement enrichie au centre du gène, H3K27me2 et H3K27me3 sont distribuées de façon uniforme tout le long de la région transcrite, l'H3K27me2 est également retrouvée sur la région flanquante. Toutes trois occupent des gènes transcrits pour la plupart à un niveau modéré ou de façon tissu-spécifique. Tandis que l'H3K27me2 et l'H3K27me3 tendent à être associées aux mêmes gènes, l'H3K27me1 affecte un sous-groupe de gènes distincts, ce qui suggère l'existence de deux voies indépendantes de répression impliquant la méthylation de l'H3K27 (**Figure I-16**).

L'H3K27me1 et me2 sont déposées sur les loci hétérochromatiques par les HMT ARABIDOPSIS TRITHORAX-RELATED PROTEIN 5 (ATXR5) et 6. Cependant, l'implication de ces protéines dans le dépôt de ces marques sur les gènes n'est pas démontrée.



**Figure I-17 : Méthylation de l'H3K9.**

L'H3K9me3 est principalement retrouvée sur l'extrémité 5' de gènes fortement exprimés. REF6 et ELF6 enlèveraient cette marque sur certains gènes de réponse aux brassinostéroïdes. La déméthylase histone IBM1 protège, elle, les gènes de la marque répressive H3K9me2.

De plus, les éléments marqués par l'H3K27me1 et me2 sont également fortement méthylés et associés à l'H3K9me2, ce qui n'est pas le cas pour les gènes ciblés, suggérant donc des mécanismes d'action ou des partenaires différents.

### La méthylation de l'H3K9

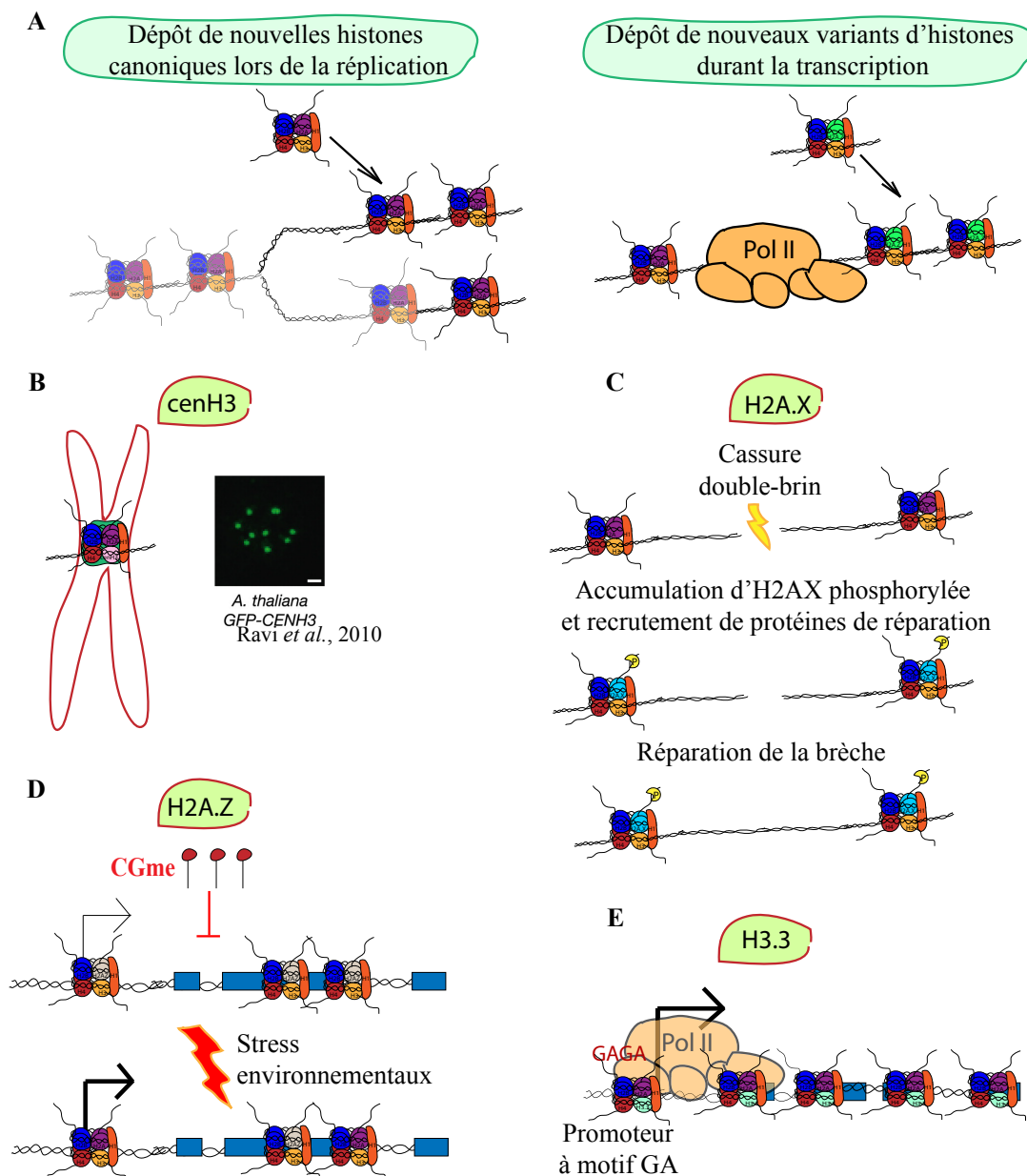
La marque H3K9me2 répressive est presque exclusivement retrouvée dans les éléments transposables en contexte sauvage. *A contrario*, l'H3K9me3 se localise exclusivement sur les gènes, et est enrichie au niveau de leur extrémité 5' (Roudier et al. 2011) (**Figure I-17**). Ces gènes cibles sont fortement exprimés. Le rôle de cette marque et son effet sur la transcription reste à déterminer.

### Les histones déméthylases

En comparaison avec d'autres modifications post-traductionnelles, la méthylation des histones est relativement stable mais peut être modulée par l'action d'histones déméthylases enlevant les groupements méthyl. Chez *Arabidopsis*, quatre gènes homologues de la *LYSINE-SPECIFIC DEMETHYLASE 1* (*LSD1*), *LSD1-LIKE 1* (*LDL1*), *LDL2*, *LDL3* et *FLOWERING LOCUS D* (*FLD*), participent à la déméthylation H3K4 (**Figure I-14**). Vingt-et-une protéines contenant un domaine Jumonji (JmjC) à activité histone déméthylase ont également été identifiées et quatre d'entre elles ont été caractérisées.

JUMONJI 15 (JMJ15)/MEE27 et JMJ14 affectent l'H3K4me1, me2 et me3 (Liu et al. 2010, Lu et al. 2010) (**Figure I-14**). De façon intéressante, LDL1, LDL2 et JMJ14, par leur action de déméthylation de l'H3K4 permettrait de limiter l'expansion d'H3K4me3 autour des promoteurs, préservant ainsi l'accès des régions voisines à la voie RdDM afin de maintenir, et une transcription correcte des gènes, et une répression des éléments transposables voisins (Greenberg et al. 2013).

EARLY FLOWERING 6 (ELF6) affecterait l'H3K9me3, comme RELATIVE OF EARLY FLOWERING 6 (REF6)/JMJ12 qui participe aussi à la déméthylation de l'H3K27me2 et me3 (Yu et al. 2008, Lu et al. 2011) (**Figures I-16 et 17**). Enfin, la protéine INCREASE IN BONSAI METHYLATION 1 (IBM1)/JMJ25 est une histone déméthylase H3K9. Elle protège les gènes de la marque H3K9me2 répressive déposée par KRYPTONITE (KYP). Le mutant *ibm1* présente l'accumulation ectopique d'H3K9me2 ainsi que de CHGme dans les régions transcrites de nombreux gènes à expression constitutive, ainsi que des anomalies développementales. Cette histone déméthylase a donc une fonction importante dans le développement correct de la plante.



**Figure I-18 : Les variants d'histones.**

**A-** Les **variants d'histones** peuvent être déposés au cours de la transcription tandis que les histones canoniques s'incorporent lors de la réplication.

**B-** **CenH3** est un variant retrouvé dans les zones centromériques, il jouerait un rôle important dans l'établissement des kinétochores et la ségrégation des chromosomes lors de la méiose et de la mitose.

**C-** Lors de cassures double-brin, l'**H2A.X** se regroupe en foci et est phosphorylé, ce qui provoque le recrutement de protéines de réparation comblant la brèche.

**D-** L'**H2A.Z** est retrouvé, dans les plantes sauvages, près du site d'initiation de transcription et dans le corps de gènes de réponse aux stress abiotiques. Lors de stress, la perte d'H2A.Z déclencherait une dérégulation de l'expression de ces gènes. Ce variant est antagoniste à la méthylation CG.

**E-** Le variant **H3.3** est localisé dans les régions transcrites de certains gènes ou au niveau de promoteurs à motif GA, il faciliterait l'accès de la polymérase II à ces promoteurs.

#### I.A.4.a.4. Les variants d'histones

La chromatine est une structure très variable et dynamique ; les histones, cœur du nucléosome, peuvent subir des modifications post-traductionnelles mais également être remplacées par des variants, isoformes non alléliques comportant quelques variations dans la séquence polypeptidique.

Le génome d'*Arabidopsis* contient huit gènes codant l'histone H4, quinze gènes pour l'histone H3, treize gènes pour l'histone H2A, onze pour l'histone H2B et enfin trois gènes codant l'histone H1. Tandis que les gènes d'histones canoniques, les plus nombreux, sont exprimés majoritairement pendant la réplication et ne contiennent pas d'introns, les variants d'histones sont produits et peuvent être incorporés durant tout le cycle cellulaire, et leur séquence nucléotidique comprend des introns (**Figure I-18A**). Chaque variant est retrouvé spécifiquement à certains loci dont il modifie la structure chromatinienne et/ou l'expression. Les variants codant pour des histones H3 et H2A sont les mieux caractérisés, comme décrit ci-après, ils ont des effets variables sur les processus cellulaires fondamentaux.

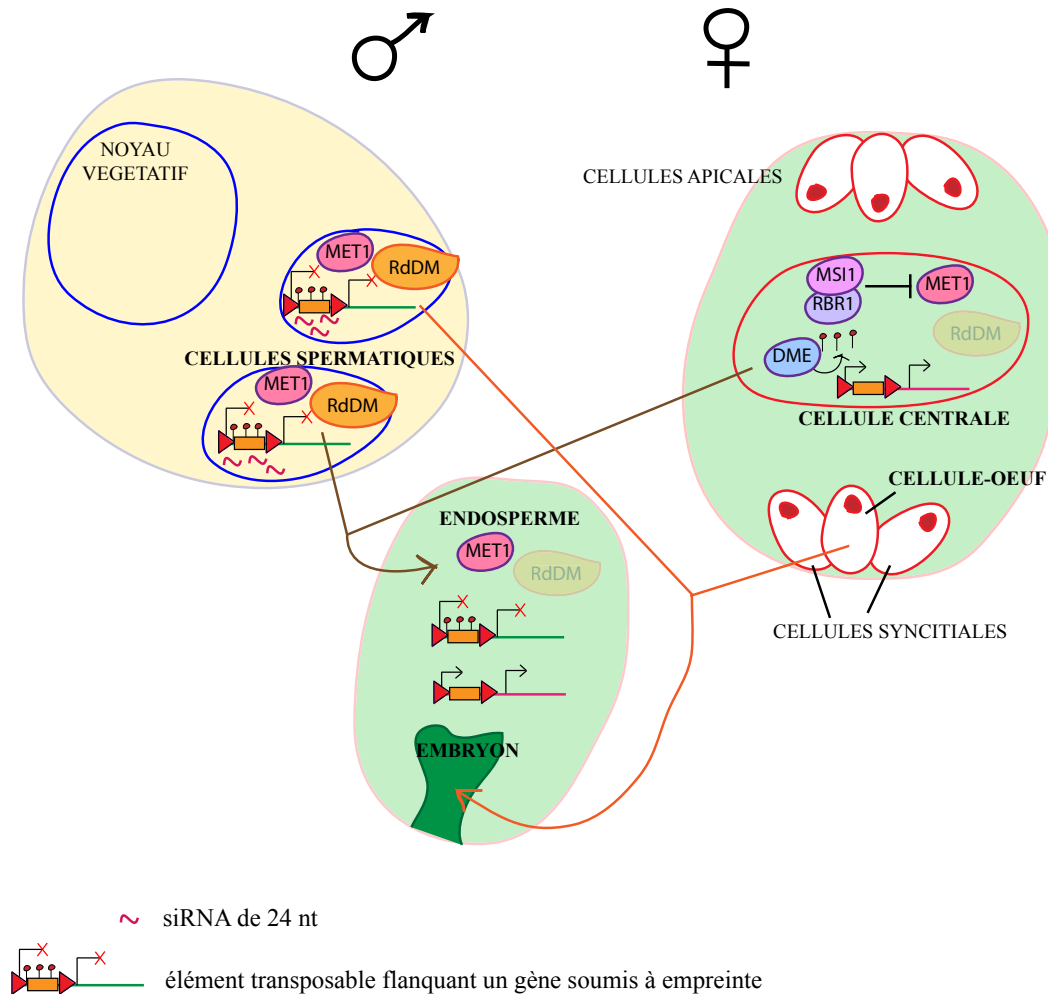
Ainsi, le variant CenH3, s'incorporant au niveau des centromères, participe à l'établissement des kinétochores et à la ségrégation des chromosomes durant la méiose et la mitose (**Figure I-18B**).

H2A.X, lui, se regroupe en foci au niveau de sites de dommages à l'ADN et est impliqué dans les voies de réparation de l'ADN (**Figure I-18C**).

Deux variants sont majoritairement retrouvés sur les gènes, H3.3 et H2A.Z.

L'H2A.Z est le variant d'histone le plus conservé, il est incorporé par un complexe homologue au complexe SWI2/SNF2 RELATED (SWR1) de la Levure. Ce variant est retrouvé au niveau des sites d'initiation de transcription et dans le corps de nombreux gènes et est anti-corrélé à la méthylation ADN (Zilberman et al. 2008) (**Figure I-18D**). Les gènes cibles d'H2A.Z sont le plus souvent modérément exprimés et la présence d'H2A.Z dans le corps des gènes permettrait de moduler leur expression en réponse à des stimuli développementaux ou environnementaux (Coleman-Derr et al. 2012b). Chez *Arabidopsis*, trois gènes codent pour ce variant, *HTA9*, *HTA11*, et *HTA8* (Redon et al. 2002, Yi et al. 2006).

L'H3.3 est, elle, localisée dans les régions transcrites, en association avec la transcription, mais également dans les promoteurs de gènes -actifs ou inactifs- portant des motifs GA, où il colocalise avec l'ARN polymérase II (**Figure I-18E**). Ces régions promotrices sont la plupart du temps dénuée de méthylation CG mais enrichies en H3K27me3.



**Figure I-19 : Implication de la voie RdDM dans la régulation des gènes soumis à empreinte parentale.**

Les gènes soumis à empreinte sont réprimés par la voie RdDM dans le gamétophyte mâle. Ils sont par contre exprimés dans la cellule-œuf, où les gènes codant des protéines de la voie RdDM sont faiblement transcrits. Après fécondation, ces gènes demeurent exprimés maternellement dans l'endosperme.

Les nucléosomes comportant ce variant sont plus accessibles, l'H3.3 pourrait donc affecter la condensation de la chromatine et ainsi faciliter l'accès de la polymérase II pour une transcription immédiate ou amorcer une activation postérieure des gènes cibles (Shu et al.).

#### ***I.A.4.b. La voie RNA-directed DNA methylation (RdDM)***

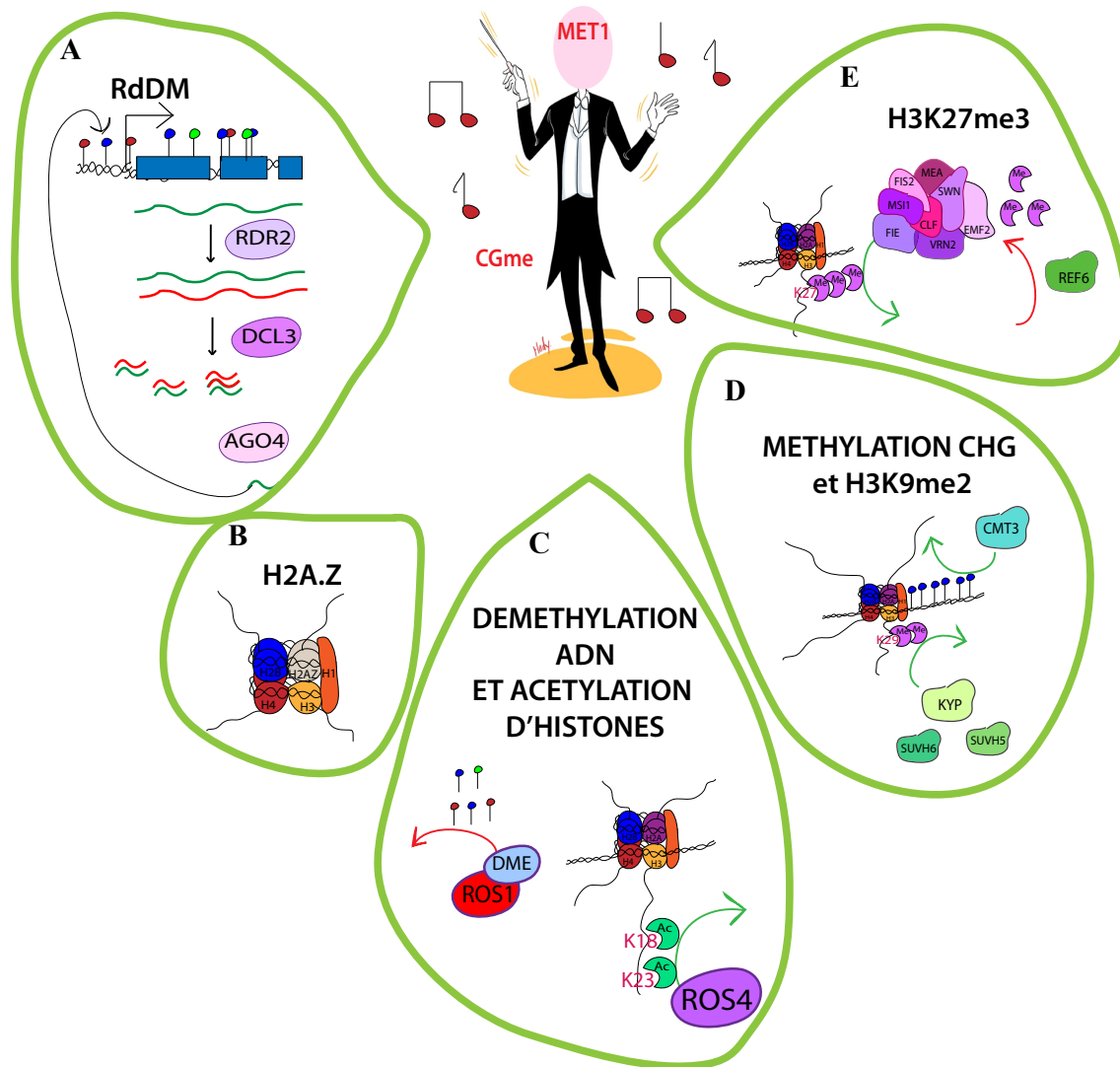
La voie RdDM implique la génération de petits ARN de 24 nt, des small interfering RNA (siRNA), initiée par les polymérases ARN IV et V spécifiques des Plantes. L'ARN polymérase ARN dépendante RDR2 produit des ARN double brin coupés ensuite par la protéine DCL3 dans le nucléole pour former des ARN simple brin de 24 nt. Ils sont ensuite stabilisés par la méthylation de leur extrémité 3' via l'action d'HEN1 puis chargés sur la protéine AGO4 jusqu'aux loci cibles afin de guider l'ADN méthyltransférase DRM2 (Kanno et al. 2011, Bologna et al. 2014). Ainsi les loci homologues des siRNAs sont méthylés *de novo* et réprimés. De nombreux cofacteurs participant à cette voie ont été identifiés récemment (Gao et al. 2010, Law et al. 2011, Eun et al. 2012, Huang et al. 2013, Zhang et al. 2013, Groth et al. 2014, Zhong et al. 2015).

Même si cette voie est fortement impliquée dans la répression des loci hétérochromatiques, elle a aussi un rôle crucial dans le contrôle de l'expression de certains gènes. En effet, de nombreux promoteurs et/ou corps de gènes sont proches ou chevauchent des séquences répétées ou des éléments transposables. La voie RdDM, en collaboration avec DDM1, cible ces répétitions et ainsi préserve l'expression appropriée du gène (Zemach et al. 2013).

De plus, la voie RdDM est impliquée dans le processus d'empreinte parentale, dans le développement ainsi que dans la défense de la plante. Ainsi, certains loci réprimés par cette voie dans les tissus somatiques sont exprimés maternellement dans l'endosperme, en conséquence de la faible expression, dans les étapes précoces du développement, des gènes impliqués dans la RdDM (**Figure I-19**) (Vu et al. 2013). Dans les cellules spermatiques du gamétophyte mâle où les glycosylases ADN ne sont pas exprimées, les éléments transposables, notamment ceux flanquant les loci soumis à empreinte, accumulent des siRNA de 24nt qui guident la reméthylation et la répression des loci flanquants (Calarco et al. 2012).

La voie RdDM ainsi que l'H3K9me2 participent à la répression des allèles paternels dans les phases d'embryogenèse précoce également, expliquant la forte prédominance du transcriptome maternel (Autran et al. 2011). Les transcrits paternels s'accumulent ensuite graduellement du stade 2-4 cellules au stade globulaire, grâce à des échanges d'histones.





**Figure I-20 : La méthylation CG, chef d'orchestre épigénétique.**

La méthylation CG joue un rôle crucial dans diverses voies épigénétiques et maintient les profils épigénétiques corrects au niveau des gènes.

**A-** Elle contrôle la voie RdDM produisant des petits ARN.

**B-** La méthylation CG dans les corps des gènes de ménage pourraient exclure le variant H2A.Z pour préserver leur expression constitutive.

**C-** Elle régule l'expression des déméthylases ADN *DME* et *ROS1*. De plus, l'acétyltransférase histone *ROS4* reconnaît les sites CG méthylés et acétylerait l'H3K18 et l'H3K23 provoquant ainsi le recrutement de *ROS1* sur ces loci.

**D-** La méthylation CG stabilise les profils de méthylation CHG et H3K9me2 sur des gènes.

**E-** Enfin, l'H3K27me3 présente sur de nombreux gènes dans les plantes sauvages se relocalise vers des loci hétérochromatiques en absence de méthylation CG.

Les variants H3.3, majoritaires dans les gamétophytes mâle et femelle, sont notamment activement enlevés après fécondation. CAF1 charge des variants H3.1 dans la chromatine de l'embryon, limitant ainsi la transmission de l'information épigénétique, notamment des modifications post-traductionnelles affectant les queues des histones parentales (Autran et al. 2011) (Ingouff et al. 2010).

Parmi les loci soumis à empreinte, des gènes impliqués dans la floraison, dans la régulation hormonale ou de la chromatine ainsi que dans la réponse aux pathogènes sont particulièrement représentés (Berger et al. 2012).

La voie RdDM régule également l'expression de gènes de réponse aux stimuli environnementaux. Ces gènes sont activés lors de stress biotiques ou abiotiques notamment par répression transitoire des gènes de la voie RdDM (Downen et al. 2012b, Yu et al. 2013).

Cette voie de silencing a donc une action prédominante dans la répression des gènes mais est également engagée dans le contrôle de l'expression des gènes, dont de nombreux gènes soumis à empreinte, nécessaires au développement de la plante ou impliqués dans la réponse aux stimuli environnementaux.

#### ***1.A.4.c. Rôle de la méthylation CG sur les gènes***

Toutes les marques épigénétiques présentées dans les paragraphes précédents participent au silencing des éléments transposables, à l'expression correcte des gènes ainsi qu'à la réplication, la réparation ou la recombinaison. L'association de diverses marques au niveau des loci permet une régulation fine de chacun, selon les tissus, les étapes du développement, les conditions environnementales.

Tandis que sur les éléments transposables, un « mille-feuille » constitué de multiples marques épigénétiques répressives permet de verrouiller leur expression et leur mobilisation, sur les gènes ont également pu être mis en évidence des liens étroits entre les marques, liens coordonnés par la méthylation CG (**Figure I-20**).

En effet, le mutant *met1*, montre, parallèlement à la perte en méthylation CG, une hyperméthylation sur les sites non-CG de nombreux loci (**Figure I-20D**). Sur une centaine de ces loci, l'apparition ectopique de méthylation CHG et CHH a été attribuée à une relocalisation stochastique de la voie RdDM ainsi qu'à la répression totale de *ROS1* et partielle de *DME* dans le mutant (Mathieu et al. 2007) (**Figure I-20A&C**).



De façon intéressante, ces deux phénomènes sont eux-mêmes liés, puisque de nombreux mutants de la voie RdDM réduisent le taux d'expression de *ROS1* (Huettel et al. 2006).

La protéine ROS4/IDM1, une HAT acétylant les résidus lysine 18 et 23 de l'histone H3 serait capable de reconnaître les sites CG méthylés, l'H3K4 non méthylée et l'H3R2 non méthylée, et notamment de guider la déméthylase ADN ROS1 vers ses cibles, assistée de la protéine ROS5 (Li et al. 2012, Qian et al. 2012, Zhao et al. 2014). La méthylation CG pourrait donc paradoxalement recruter l'HAT et activer ainsi la déméthylation de certains loci.

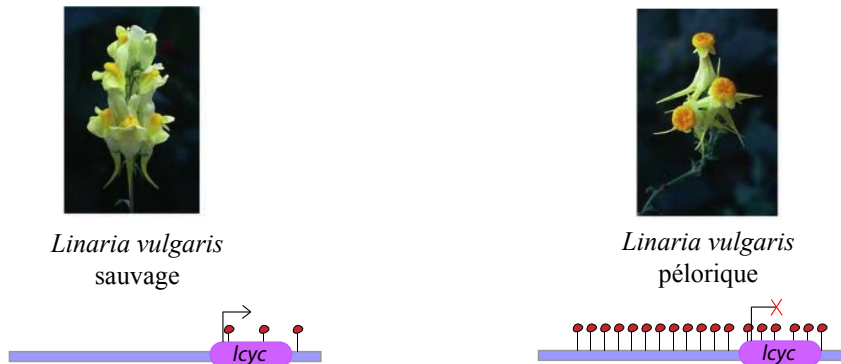
Cependant, la cause de l'hyperméthylation CHG sur quelques autres milliers de loci, n'avait, elle, pas été élucidée, au début de ma thèse. Parallèlement à cela, les analyses cytologiques de plantes *met1* ont mises en évidence une relocalisation d'H3K9me2 des chromocentres riches en éléments transposables vers des loci euchromatiques (**Figure I-20D**). Ces données suggéraient une association de la méthylation CHG et de la H3K9me2 sur les gènes dans *met1*, de même que sur les éléments transposables en contexte sauvage (Mathieu et al. 2005b).

A l'opposé, l'H3K27me3, enrichie au niveau des gènes dans les plantes sauvages, se relocalise spécifiquement vers certains loci hétérochromatiques dans le mutant *met1*, en corrélation avec leur réactivation (**Figure I-20E**). Un gain en H3K4me2 est également observé au niveau des répétitions centromériques et des éléments transposables réactivés dans le mutant hypométhylé (Mathieu et al. 2005b).

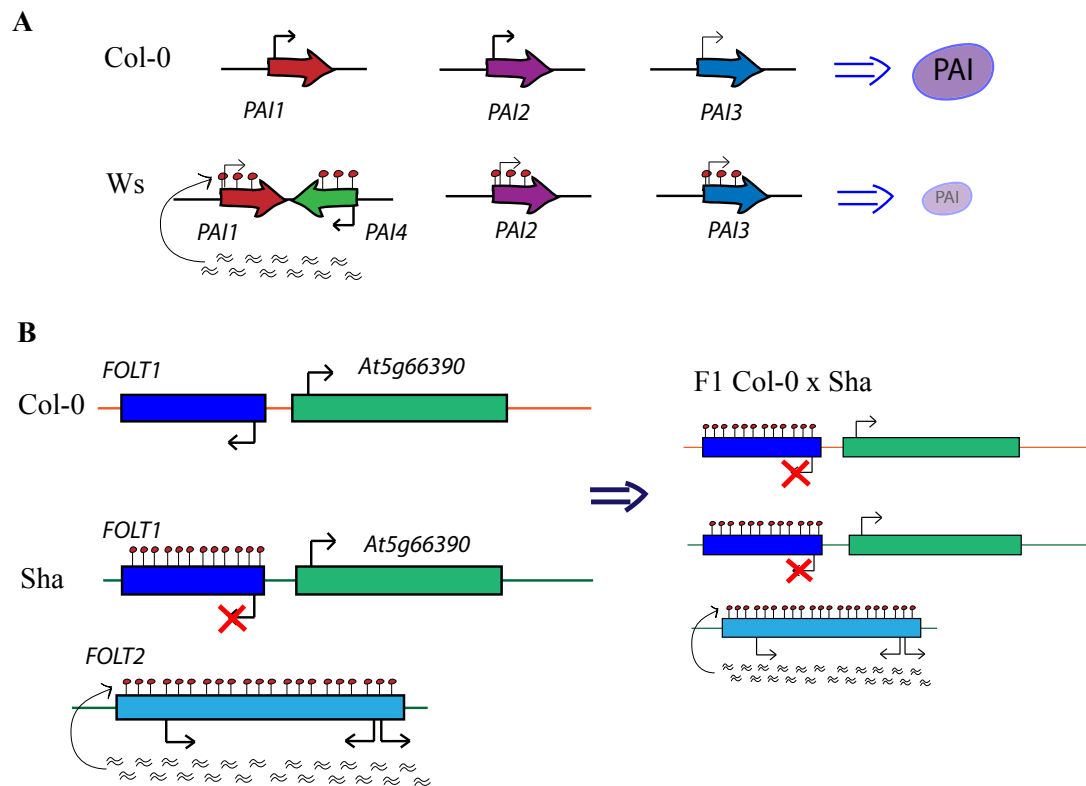
La perte de la protéine de remodelage de la chromatine DDM1 provoque également une hyperméthylation au niveau de sites CHG dans le corps de nombreux gènes, comportant tous de la méthylation CG dans les plantes sauvages (O'Neill et al. 1998, Stroud et al. 2013). Contrairement à celle observée dans *met1* surtout localisée près de l'extrémité 5', l'apparition ectopique de CHGme tend à affecter majoritairement l'extrémité 3' des gènes. L'origine de cette hyperméthylation est à ce jour inconnue.

L'H2A.Z est, lui, anticorrélé à la méthylation ADN (**Figure I-20B**). La méthylation présente dans le corps des gènes de ménage pourrait participer à exclure ce variant d'histone de ces gènes afin de préserver leur expression constitutive (Coleman-Derr et al. 2012b).

Ainsi, la méthylation CG agit comme un chef d'orchestre sur les gènes, elle régule l'expression et le recrutement de facteurs épigénétiques clés impliqués dans le dépôt de marques histones mais aussi la déméthylation de l'ADN et la voie RdDM, et contrôle ainsi la structure chromatinienne et la transcription des gènes.



**Figure I-21 : Epiallèle spontané chez *Linaria vulgaris*.** Des fleurs à la symétrie anormale apparaissent de façon spontanée. Ce phénotype “pélorique” est dû à l’hyperméthylation du locus *lcyc* provoquant le silencing du gène contrôlant la symétrie florale.



**Figure I-22 : Epiallèles spontanés observés chez *A. thaliana*.**

**A- Variation épigénétique naturelle au niveau du locus *PAI*.** Les plantes Col-0 possèdent trois gènes *PAI* non liés, codant pour une enzyme de biosynthèse du tryptophane. Chez WS, une duplication d’un gène *PAI* a généré une répétition inversée des gènes *PAI1* et du gène nouvellement formé *PAI4*. Cette répétition est ciblée par la voie RdDM provoquant la méthylation des quatre gènes *PAI*. Ainsi, l’enzyme n’est produite qu’en faible quantité.

**B- Variation épigénétique naturelle au niveau du gène *FOL1*.** Col-0 contient un gène *FOL1* tandis que plusieurs accessions dont Sha possèdent également un gène *FOL2*, comprenant plusieurs séquences répétées. La transcription de ces répétitions génère des petits ARN provoquant la méthylation des deux gènes *FOL1*. Le croisement entre Sha et Col-0 cause, dans les plantes F1, la méthylation du gène *FOL1* issu du parent Col-0 et ainsi une déficience en folate responsable d’un taux important de stérilité.

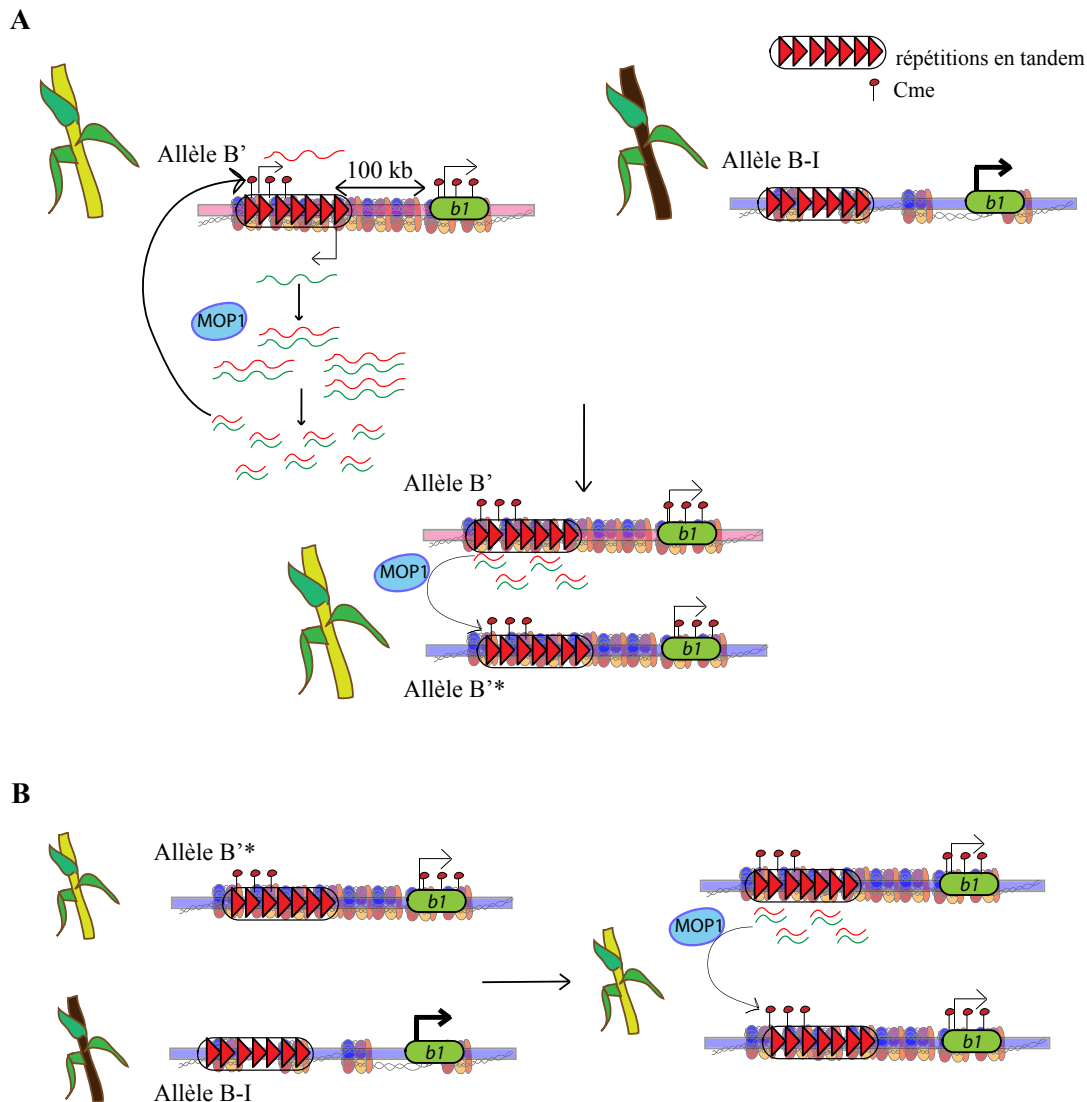
## I.B. Stabilité de la méthylation ADN au cours des générations

### I.B.1. Les épiallèles naturels

Lorsque Linnaeus, en 1744, découvre un spécimen particulier de *Linaria vulgaris*, qu'il nomme « Péloria » (« monstre » en grec), présentant de façon spontanée des fleurs à la symétrie anormale, il ne pouvait se douter de la portée de cette découverte sur le domaine de l'épigénétique et la transmission de l'information génétique (**Figure I-21**). En effet, 250 ans plus tard, Cubas *et al.*, démontrèrent que ce phénomène spontané est, non pas lié à une mutation de la séquence d'ADN elle-même, mais à l'hyperméthylation du gène *Lcyc* (Cubas *et al.* 1999). Une réversion du phénotype peut survenir, à une fréquence faible, sur un embranchement de la plante, en corrélation avec une restauration du taux de méthylation du gène. Cette découverte a ainsi démontré que de nouveaux allèles épigénétiques ou épiallèles naturels peuvent se former de façon spontanée et se transmettre ensuite au cours des générations.

D'autres exemples d'épiallèles naturels spontanés transmis à la méiose ont ensuite été décrits, ils sont fréquemment établis suite à des réarrangements chromosomiques, comme le cas du locus *PAI*, chez *Arabidopsis thaliana* (**Figure I-22A**). Ce locus code pour la famille de gènes phosphoribosylanthranilate isomérase. Ces gènes sont au nombre de trois dans l'écotype Col-0 (*PAI1*, 2 et 3) mais un quatrième membre, *PAI4* est également retrouvé dans certains écotypes tels que Wassilewskija (Ws) (Bender *et al.* 1995). Dans cet écotype, il est localisé face au gène *PAI1*, la grande répétition inverse ainsi formée est ciblée par la voie RdDM. Le gène *PAI1* est en conséquence moins transcrit que dans Col-0. De plus, ceci provoque également la méthylation des deux gènes non liés, *PAI2* et 3, et par ce biais, la répression du gène *PAI2*, ce qui engendre une réduction de l'activité enzymatique PAI. La méthylation des deux gènes *PAI2* et 3 peut être transmise à la méiose, et le croisement avec des plantes Col-0 provoque le ciblage des allèles *PAI* non méthylés par la RdDM. Ainsi, ceci suggère que des séquences homologues peuvent communiquer à travers le génome et transmettre leur état épigénétique.

Un phénomène similaire a été rapporté au niveau du gène *FOLate Transporter 1 (FOLT1)* dupliqué dans certaines accessions d'*A. thaliana* (**Figure I-22B**). La duplication génère le gène *FOLT2* incluant des séquences tronquées réarrangées. Cette structure réarrangée est méthylée et induit la méthylation *de novo* en *trans* des séquences homologues présentes au niveau du gène original *FOLT1*, provoquant son silencing. Le croisement de ces accessions à des plantes Col-0 ne portant, elles, que le gène *FOLT1* génère dans la descendance des plantes n'exprimant ni *FOLT1* ni *FOLT2*.



**Figure I-23 : Processus de paramutation observé au niveau du locus *booster 1* (*b1*) codant pour l'anthocyanine, un pigment de la tige.**

**A-** Chez les plantes portant l'allèle paramutagène *B'*, le locus est méthylé, le gène *b1* est peu exprimé, les plantes présentent donc des tiges claires. Cette structure condensée est causée par la transcription des deux brins des répétitions situées à 100 kb en amont du gène *b1*. Elle permet la formation d'ARN double brin (db) via l'action de la protéine MOP1, ensuite coupés en petits ARNdb. Les plantes *B-I*, au contraire, ont, elles, des tiges foncées puisque le gène *b1* est fortement exprimé, le locus est peu méthylé.

**B-** Lorsque des plantes *B'* sont croisées à des plantes *B-I*, l'allèle *B-I* est paramuté et le locus *b1* est peu exprimé, les plantes issues du croisement ont donc des tiges claires. De plus, cet allèle, maintenant nommé *B'\**, est capable à son tour de paramuter un allèle *B-I*.

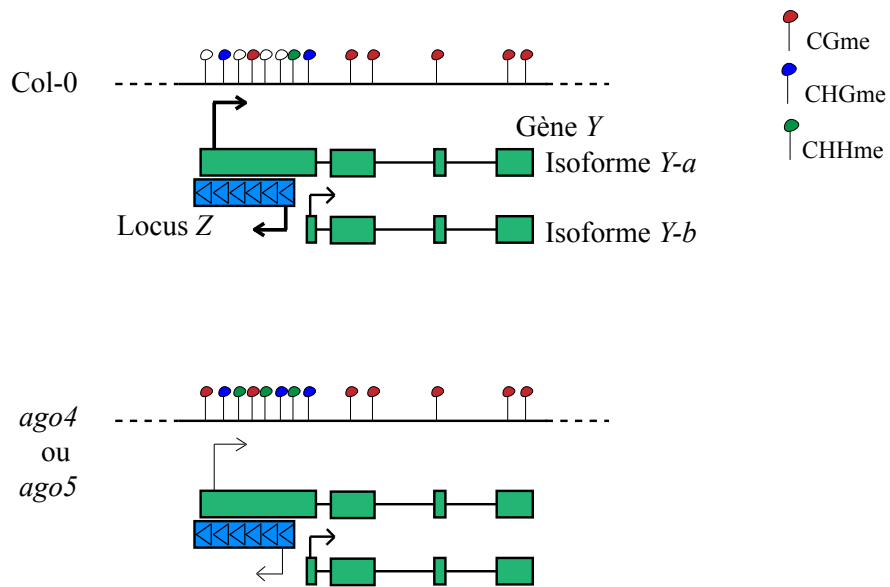
Ces plantes déficientes en folate ont une fertilité réduite. Ainsi, cette étude a mis en évidence, pour la première fois, un impact délétère d'épiallèles naturels sur le phénotype de la plante, ici une incompatibilité génétique (Durand et al. 2012).

Ce mécanisme de méthylation en *trans* a été davantage disséqué chez le Maïs, notamment au niveau du locus *booster 1* (« *b1* ») (**Figure I-23A**) (Coe 1959). Ce locus encode un facteur de transcription régulant des gènes impliqués dans la biosynthèse de pigments flavonoïdes. L'allèle *B-I*, fortement transcrit, produit des grains foncés tandis que l'allèle *B'*, faiblement transcrit, des grains clairs. Leur niveau d'expression est inversement corrélé au taux de méthylation d'une région située en amont du locus, comportant sept répétitions directes. Le croisement de plantes *B-I* à des plantes *B'* engendre une hyperméthylation et une répression de l'allèle *B-I* (**Figure I-23B**). Cet allèle paramuté, dénommé *B'\**, est ensuite capable d'agir comme l'allèle *B'* et de méthyler et réprimer l'allèle paramutable *B-I*. De plus, sa méthylation est maintenue et ne nécessite plus la présence de l'allèle *B'* qui l'a initiée. Aucune réversion de *B'\** en *B-I* n'a été observée jusqu'alors, cet épiallèle est donc particulièrement stable au cours des générations.

Ce phénomène de paramutation a été observé sur d'autres loci chez le Maïs tels que *red1* (*r1*), *pericarp color1* (*p1*) ou *purple plant1* (*pl1*) et met en jeu des facteurs de la voie RdDM, Mediator Of Paramutation 1 (MOP1) et MOP2, orthologues des protéines RDR2 et NRPD2, sous-unité commune aux polymérases ARN Pol IV et Pol V. MOP1 participerait à l'amplification du pool d'ARN issus de la transcription des deux brins des répétitions en tandem présentes en amont du locus *B'* (**Figure I-23A**). Cette étape permettrait d'atteindre un niveau seuil de transcrits provoquant en *trans*, possiblement par le biais de la génération de siRNA, l'établissement puis le maintien d'un état chromatinien répressif (Alleman et al. 2006, Arteaga-Vazquez et al. 2010). D'autres protéines accessoires, dont la protéine de remodelage de la chromatine RMR1, mais aussi RMR2 et CBBP participeraient également au processus, sur certains des loci analysés (Hale et al. 2007, Brzeska et al. 2010, Barbour et al. 2012).

2000 loci pourraient ainsi être paramutables chez le Maïs. La majorité d'entre eux sont méthylés mais l'impact de cette méthylation ADN sur le silencing des loci n'est pas certain. De plus, des éléments transposables ou des séquences répétées sont fréquemment retrouvés en amont des loci paramutables et seraient responsables de la communication en *trans* des séquences homologues. Enfin, des chercheurs ont également réalisé un screen afin de rechercher des variants naturels de méthylation ADN, associés à des changements transcriptionnels. Cela a mené à la découverte d'une famille d'éléments transposables jusqu'alors non caractérisée, les rétrotransposons non-LTR SADHU.





**Figure I-24 : Variation épigénétique naturelle chez *Arabidopsis thaliana*.**

### Modèle de locus hypervariable

La méthylation de ces loci est très variable selon les fonds génétiques. Elle affecte l'expression du locus mais également du gène adjacent. En effet, le gène voisin possède deux isoformes. Le transcrit le plus long, dont l'extrémité 5' chevauche le locus hypervariable, est réprimé par la méthylation de la zone chevauchante. La cause de l'instabilité de méthylation à ces loci est à ce jour inconnue.

Un grand nombre d'éléments pleine longueur de cette famille est fortement exprimé et peu méthylé dans la lignée Col mais réprimé dans de nombreuses autres accessions (Rangwala et al. 2006).

Ainsi, des épiallèles spontanés ont été observés au niveau de gènes mais également d'éléments transposables, ces variants pouvant être transmis aux générations suivantes et affecter l'expression des loci touchés. De plus, l'étude de ces épiallèles a permis de mettre en évidence le phénomène de paramutation, communication en *trans* de loci homologues.

## **I.B.2. Variabilité épigénétiques naturelle**

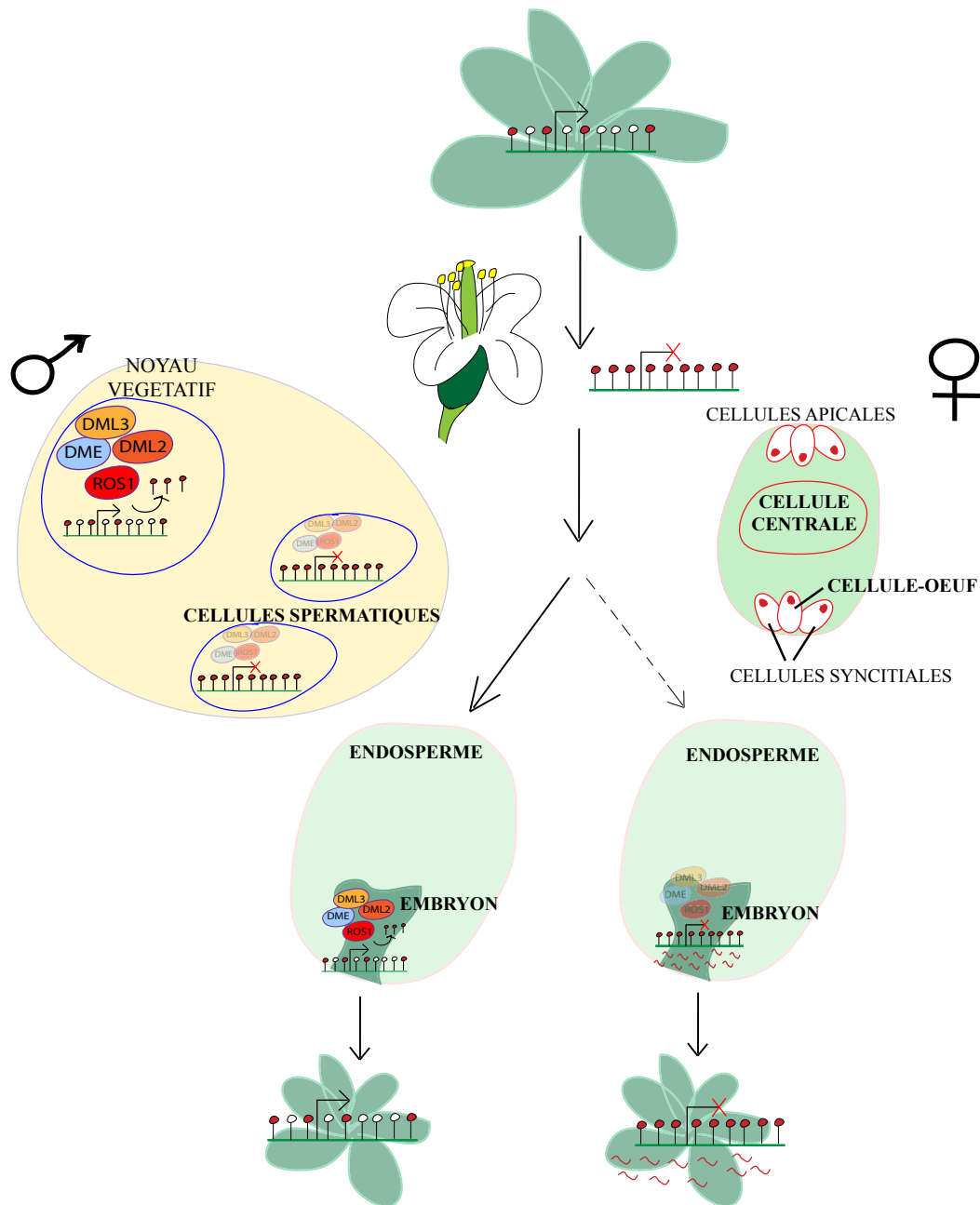
### ***I.B.2.a. Profils épigénétiques des populations naturelles***

L'analyse des profils de méthylation dans des populations naturelles d'*Arabidopsis thaliana* a apporté des informations substantielles sur la fréquence de formation des épiallèles et leur stabilité.

Une première étude, comparant les écotypes Col-0 et Ler-0, a démontré qu'alors que les éléments transposables demeurent très méthylés dans les deux accessions, les gènes présentent un taux significatif de polymorphisme épigénétique (Vaughn et al. 2007a). Le séquençage à haut-débit des profils de méthylation ADN de populations de plantes Col-0 générées par autofécondation d'un ancêtre unique sur 30 générations a confirmé cette observation sur les sites CG. La méthylation CG serait donc transmise plus stablement sur les éléments transposables, pérennisant ainsi leur répression (Becker et al. 2011, Schmitz et al. 2011). De plus, au cours des générations, tandis que de nouvelles régions différenciellement méthylées (DMR) en contexte CG apparaissent, d'autres disparaissent. Ces changements surviennent quelques fois en une génération seulement, et peuvent affecter l'expression des loci ciblés. Cette variabilité épigénétique se révèle plus importante entre les descendants qu'entre l'ancêtre commun et chacun des descendants. De plus, la fréquence de ces épimutations est considérablement supérieure au taux de mutations génétiques évalué dans ces lignées.

De façon intéressante, alors que la formation d'épiallèles spontanés demeure un phénomène rare, certaines positions se révèlent particulièrement enclines à des changements de méthylation ADN. Ces épiallèles sont donc qualifiés de métastables ou hypervariables.

Un des épiallèles métastables identifiés dans ces analyses haut-débit, localisé au niveau du locus *MRD1*, a été étudié ensuite par Havecker *et al.* (**Figure I-24**) (Havecker et al. 2012).



**Figure I-25 : Loci hypervariables chez *A. thaliana***

Les loci hypervariables tendent à être des cibles des déméthylases ADN, ou à être fortement méthylés dans les inflorescences sauvages. Dans les cellules spermatiques, les déméthylases ADN sont faiblement exprimées, et ces loci sont hyperméthylés. Cette « préméthylation » pourrait être héritée dans l'embryon après fécondation par l'accumulation de siRNAs de 24 nt, qui contrediraient l'activité de ROS1.

Sa méthylation affecte l'expression du gène *YAZ* situé en orientation inverse. En effet, le locus hypervariable chevauche l'extrémité 5' du transcrit le plus long de *YAZ*. La méthylation de la zone chevauchante dans certains fonds génétiques engendre une réduction de l'expression de l'isoforme longue du gène *YAZ* ainsi que du locus *MRDI*. Un autre locus, dénommé *RITA*, se comporterait de façon similaire. Enfin, un autre locus hypervariable, *Qua-Quine Starch (QQS)* a été étudié en 2013 (Silveira et al. 2013). Le gène *QQS* comporte des répétitions en tandem fortement méthylées dans son promoteur et sa région 5'. De nombreux variants épialléliques ont été découverts dans des stocks Col-0 de laboratoire ainsi que lors de l'analyse de différentes accessions et populations sauvages. Comme sur *RITA* et *MRDI*, le taux de méthylation ADN au niveau des répétitions est inversement corrélé à l'expression du gène *QQS*. Toutefois, la cause de l'hypervariabilité des profils épigénétiques sur ces loci demeure floue.

Parmi les loci hypervariables présentant une hyperméthylation sporadique certains sont des cibles de ROS1, DML2 et DML3, ces déméthylases ADN pourraient donc jouer un rôle dans la régulation des profils de méthylation de ces loci. D'autres sont hyperméthylés dans les tissus floraux ainsi que dans les cellules spermatiques de plantes sauvages où les déméthylases ADN sont faiblement exprimées (**Figure I-25**) (Calarco et al. 2012). Il est possible que la méthylation des loci dans ces cellules soit transmise à une certaine fréquence à la lignée germinale. Après restauration de l'activité des déméthylases, l'hyperméthylation pourrait être maintenue dans l'embryon, via l'action de la voie RdDM (**Figure I-25**). Ainsi, ces données donnent à penser que les déméthylases ADN et la voie RdDM participent à l'hypervariabilité épigénétique observée sur certains loci.

Enfin, parallèlement à ce phénomène, une étude récente des profils de méthylation chez 152 accessions naturelles a permis de déterminer qu'environ 500 loci, cibles de la RdDM sont, eux, invariablement méthylés. Ces loci incluent autant de gènes que d'éléments transposables et sont spécifiquement exprimés dans le pollen (Schmitz et al. 2013). Leur expression dans ce type cellulaire permettrait de produire des petits ARN. Ceux-ci, transmis à la descendance, participeraient à la répression de ces loci hypervariables dans les tissus végétatifs. Cette étude a également permis de mettre en évidence un lien entre des variants génétiques, locaux ou distants, et l'apparition de certaines DMRs affectant les trois contextes, comme observé précédemment notamment sur le locus *PAI* (Bender et al. 1995).

Ainsi, même si les profils de méthylation d'*Arabidopsis thaliana* sont très stables, ils peuvent être affectés au cours des générations par des épimutations, parfois associées à des variants de séquence.

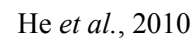


La fréquence de ces épimutations est très variable selon les régions génomiques, et pourrait dépendre d'acteurs impliqués dans l'établissement et la maintenance des profils de méthylation.

### ***1.B.2.b Impact de la formation d'hybrides sur les profils épigénétiques***

Depuis les années 1930 et le début de la Grande Dépression affectant de plein fouet l'économie mondiale, la création d'hybrides inter- ou intraspécifiques initiée sur le Maïs puis sur la majorité des espèces cultivées, s'est développée exponentiellement pour améliorer la qualité et le rendement de ces cultures. En effet, les scientifiques avaient pu observer, lors du croisement de deux lignées pures, que les hybrides qui en étaient issus étaient plus vigoureux que leurs parents. Ce phénomène est dénommé « vigueur hybride » ou « hétérosis ». Etant donné l'intérêt économique majeur, de nombreux chercheurs ont tenté de percer à jour le secret de la vigueur hybride. Ces recherches ont permis notamment de mettre en évidence le rôle crucial de l'épigénétique dans ce phénomène et de progresser dans l'étude de la stabilité des marques épigénétiques.

Les premières analyses ont concerné majoritairement des hybrides interspécifiques. Elles ont dévoilé un défaut d'accumulation des siRNA associé à la dérégulation d'éléments transposables, ceci engendrant une instabilité génomique et une réduction de la fertilité (Comai et al. 2000, Chen et al. 2008). L'étude d'hybrides *A. thaliana* x *A. arenosa* a également permis de constater que les changements dans l'accumulation de siRNA étaient particulièrement importants en F1 sur les zones où le taux de ces petits ARN était fortement divergent entre les progéniteurs. De plus, ces changements étaient maintenus de façon stable dans les générations, préservant la stabilité du génome après hybridation. Les microARN (miARN), s'accumulant eux de façon similaire dans les deux parents, sont exprimés de façon non-additive dans les hybrides, dégradant les ARNm de nombreux gènes, et médieraient donc de la diversité génique. Des gènes soumis à empreinte ainsi que des gènes homéologues sont également dérégulés dans les polyploïdes, et des rétroéléments déméthylés (Lee et al. 2001, Madlung et al. 2002, Bushell et al. 2003, Wang et al. 2004, Madlung et al. 2005, Josefsson et al. 2006b, Wang et al. 2006a, Wang et al. 2006b). Des premières analyses de méthylation par digestion enzymatique dans des hybrides ont permis de démontrer qu'au delà de la déméthylation d'éléments transposables, les profils de méthylation sont globalement fortement affectés (Jin et al. 2008). Il a été démontré ensuite que dans les plantes F1 du croisement d'*A. thaliana* et *A. lyrata*, 70% des loci affectés présentent une hyperméthylation sur les cytosines en contexte CCG majoritairement, mais aussi CG.



**A-** Phénotypes des hybrides Nipponbare x 93-11. Nipp. = Nipponbare

**B-** Variations additives et non-additives affectant l'expression et les modifications épigénétiques chez les hybrides.

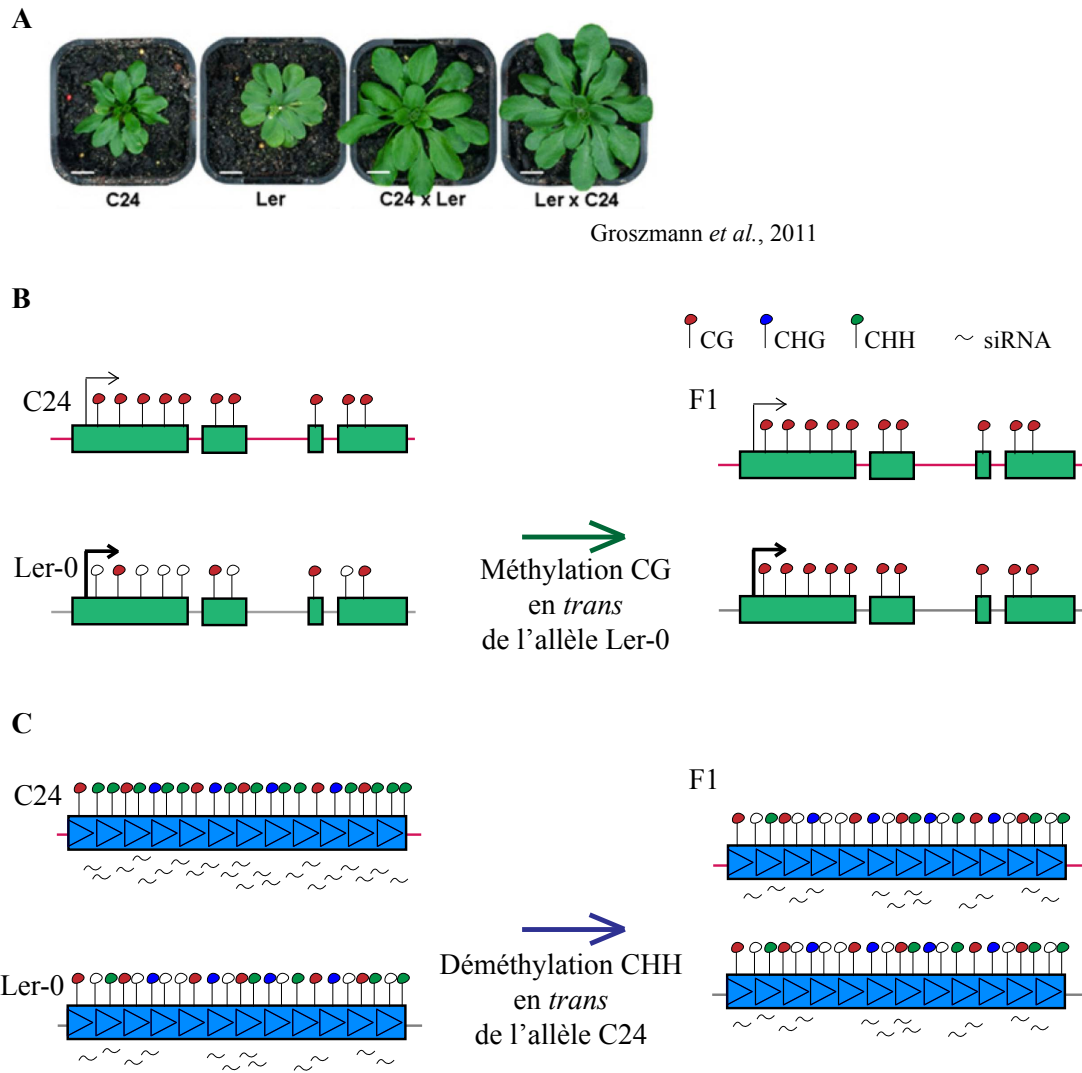
**C-** Expression différentielle des clusters de siRNA dans les hybrides F1 par rapport à la MPV (valeur moyenne des parents).

De plus, alors que l'hyperméthylation CG est stablement transmise aux générations suivantes, les sites CCG sont le plus souvent reméthylés au taux sauvage dès la seconde génération (Beaulieu et al. 2009).

Le plus souvent, le degré et la fréquence d'hétérosis corrélient avec la distance génétique des parents. Cependant, une forte vigueur hybride est notée dans certains croisements intra-spécifiques (**Figure I-26A**). Ainsi, l'étude d'hybrides réciproques issus de deux sous-espèces de Riz révèle des changements de méthylation majoritairement au niveau d'éléments transposables, ainsi qu'une expression non-additive sur de nombreux loci, pas nécessairement corrélée aux changements de méthylation ADN (**Figure I-26B**). Parmi ces loci, de nombreux gènes codant pour des fonctions métaboliques sont sur-exprimés dans les deux hybrides réciproques par rapport aux parents et pourraient donc participer à leur vigueur hybride (He et al. 2010). Enfin, comme notée dans les croisements inter-espèces, l'accumulation des petits ARN est réduite dans ces hybrides de Riz à un grand nombre de loci, quelque soit leur taux d'accumulation dans les deux parents (**Figure I-26C**).

Des observations similaires ont ensuite pu être réalisées chez *Arabidopsis thaliana*, lors du croisement des écotypes Ler-0 et C24 (**Figure I-27A**) (Groszmann et al. 2011b, Greaves et al. 2012b, Shen et al. 2012). De façon concordante aux données obtenues dans le croisement *A. lyrata* x *A. thaliana*, la méthylation CG tend à augmenter dans l'hybride, la méthylation CHH varie également fortement. Le plus souvent, ces changements ont lieu au niveau de loci différenciellement méthylés entre les deux parents, et sont pour certains corrélés à des variations transcriptionnelles (Groszmann et al. 2011b, Greaves et al. 2012b, Shen et al. 2012). Lors de la rencontre des deux génomes, le profil de méthylation de l'un des parents pourrait être transféré à l'autre par un phénomène nommé méthylation ou déméthylation chromosomique en *trans* (**Figure I-27B&C**). Une baisse d'accumulation des siRNA de 24 nt a également été observée dans les hybrides, elle touche particulièrement les zones où le taux de siRNA est très différent entre les parents et est associée à une hypométhylation CHH (Groszmann et al. 2011b, Greaves et al. 2012b). De plus, le croisement de plantes sauvages avec des plantes mutées pour le gène *HEN1*, codant une méthylase ARN impliquée dans la stabilisation des petits ARN (siRNA et miARN), génèrent des plantes de taille réduite et sans vigueur végétative. Cela suggère que la réduction en petits ARN dans les hybrides C24 x Ler n'est pas seule responsable du phénotype de vigueur hybride (Shen et al. 2012).





**Figure I-27 : Impact du croisement C24 x Ler-0 sur les taux de méthylation et de siRNA.**

#### A- Phénotypes des plantes F1 C24 x Ler-0

**B et C- Schéma de méthylation CG (B) ou déméthylation CHH (C) en *trans* observées lors du croisement de plantes C24 et Ler-0.** Lors de la rencontre des deux allèles, le profil de méthylation de l'un des parents est transféré à l'autre. Ceci est observé sur des loci dont le taux de méthylation est très différent entre les plantes parentes, et peut affecter leur expression.

De plus, le phénomène de déméthylation CHH est liée à une accumulation différentielle de siRNA de 24 nt (C).

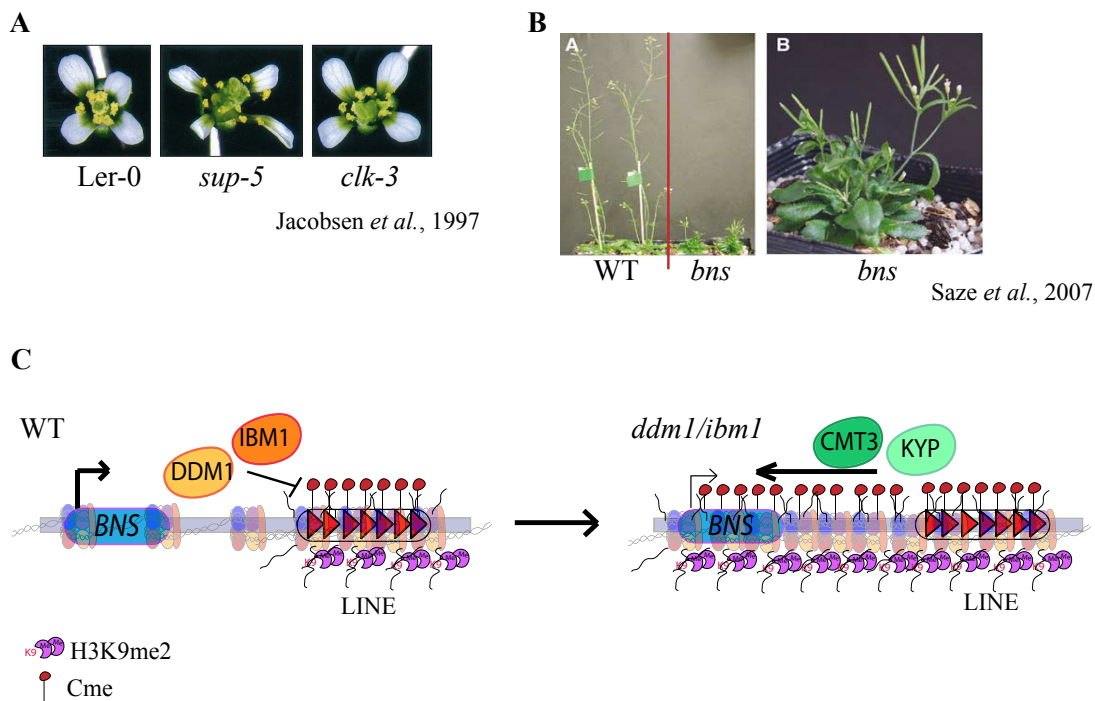
Enfin, une étude récente d'hybrides créés à partir de croisements entre accessions d'*A. thaliana* révèle que des variations épigénétiques au niveau d'éléments transposables peut affecter l'empreinte parentale et ainsi fournir de nouveaux profils d'expression dans les graines (Pignatta et al. 2014).

Ainsi, l'analyse d'hybrides a démontré que la rencontre d'épigénomes différents impacte sur la transcription, les profils de méthylation ainsi que l'accumulation de petits ARN, ces modifications épigénétiques participant au phénomène de vigueur hybride et pouvant influencer sur le processus d'empreinte parentale.

### I.B.3. Épiallèles induits dans des mutants hypométhylés

Afin d'étudier l'importance de la méthylation ADN et les acteurs impliqués dans les mécanismes de dépôt et de maintenance des profils de méthylation, des mutants hypométhylés tels que *met1* ou *ddm1* ont été analysés. Dans ces mutants, l'observation de certaines anomalies phénotypiques, pouvant rappeler le phénotype retrouvé dans des mutants déjà connus, ont permis d'identifier de nouveaux épiallèles. Parmi eux, l'épiallèle *clark kent* (*clk*), provoquant l'apparition d'étamines surnuméraires et de carpelles mal fusionnés est causé par l'hyperméthylation de répétitions CT entourant le site d'initiation de la transcription du gène *SUPERMAN* (*SUP*), ce qui provoque sa répression (**Figure I-28A**) (Jacobsen et al. 1997). Ce phénotype est retrouvé de façon sporadique après plusieurs autofécondations successives des mutants *met1* et *ddm1*. De même, l'hyperméthylation de répétitions CT dans des régions introniques du gène *AGAMOUS* (*AG*) dans une lignée portant une construction antisens du gène *MET1* est responsable de la conversion des étamines en pétales, comme dans le mutant *ag* (Bowman et al. 1989, Jacobsen et al. 2000b).

Certaines plantes *ddm1* présentent également de courtes tiges et inflorescences. Ce phénotype est associé à l'hyperméthylation du gène *BONSAI* (*BNS*) codant une protéine similaire à APC13, sous-unité du complexe Anaphase Promoting Complex (APC) (**Figure I-28B**) (Saze et al. 2007a). Un élément tronqué de type LINE est présent dans la séquence du gène et fortement méthylé dans les plantes Col-0 (**Figure I-28C**). Dans *ddm1*, la méthylation du LINE s'étend sur le gène voisin *BNS*, surtout au niveau de sites non-CG. Ceci provoque sa répression et des anomalies de croissance de la tige et d'élongation internodale.



**Figure I-28 : Epiallèles induits en laboratoire.**

**A- Phénotypes floraux de plantes présentant l'épiallèle *sup*.**

**B- Phénotypes de plantes portant l'épiallèle *bns*.**

**C- Hyperméthylation du gène *BNS* dans les mutants *ddm1* ou *ibm1*.** Un élément LINE, présent en aval du gène *BNS*, est fortement méthyle dans les plantes sauvages et enrichi en H3K9me2. Le gène *BNS* est protégé grâce à la protéine de remodelage de chromatine DDM1 et l'histone déméthylase IBM1. Dans des mutants *ibm1* ou *ddm1*, la méthylation du LINE s'étend sur le gène provoquant sa répression.

Un screen de plantes mutagénisées présentant une hyperméthylation au niveau du gène *BNS* a permis d'identifier le gène *IBM1*, cette histone déméthylase H3K9me2 protégerait ce gène de cette méthylation ectopique délétère (**Figure I-28C**) (Saze et al. 2008).

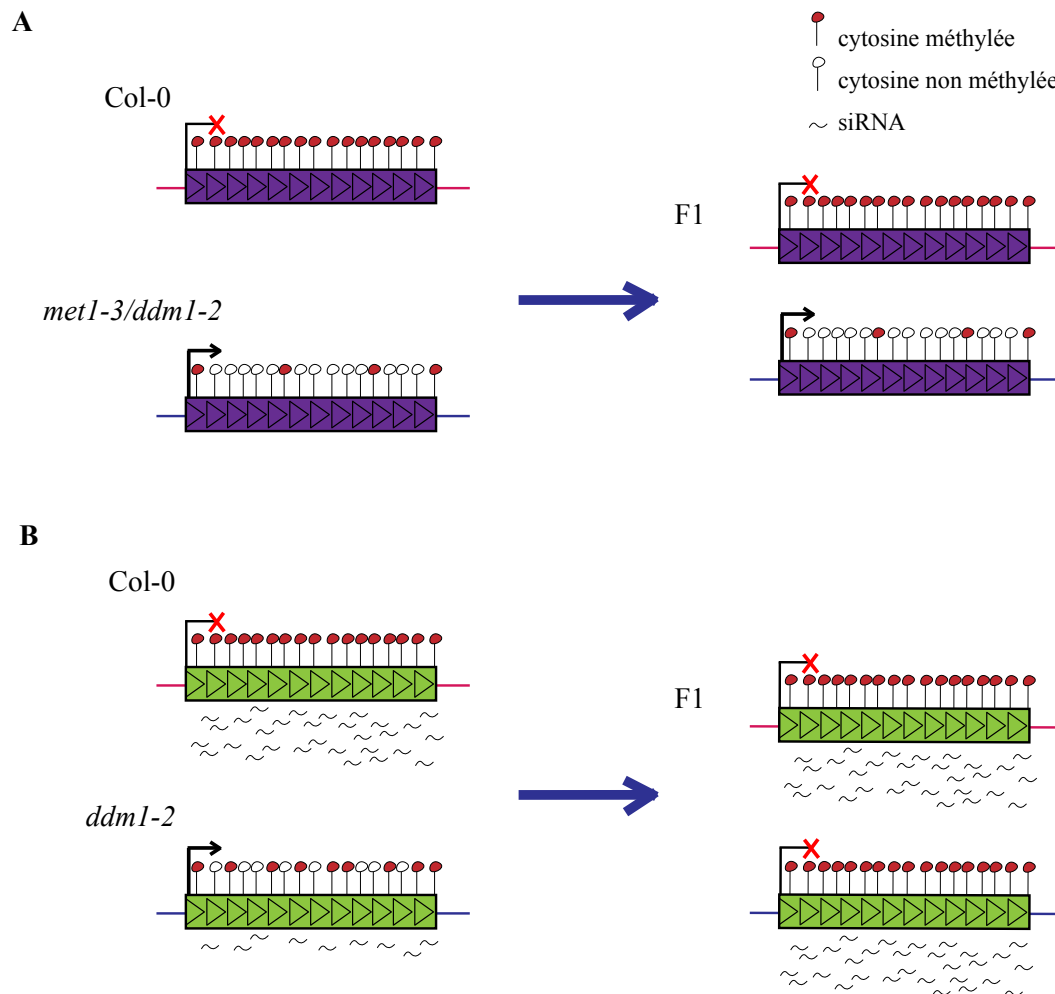
Nous pouvons aussi citer l'épiallèle *fwa*, menant à un retard de floraison, observé dans *met1* et dans les générations tardives de *ddm1* (**Figure I-6B**) (Kakutani 1997). La déméthylation de l'élément SINE présent dans le promoteur du gène est responsable de l'expression ectopique du gène dans cet épiallèle (Soppe et al. 2000a).

Ainsi, les épiallèles identifiés dans ces mutants hypométhylés sont le plus souvent liés à l'hypo- ou l'hyperméthylation ectopique de gènes voisins d'éléments transposables ou de séquences répétées. Le maintien de profils de méthylation corrects et la protection des gènes contre l'extension de l'hétérochromatine méthylée sont par conséquent cruciaux pour le développement correct de la plante.

#### I.B.4. Rétrocroisement de mutants hypométhylés

Afin de progresser dans la compréhension de la transmission des profils épigénétiques, certains éléments transposables ont spécifiquement été analysés dans des plantes issues du croisement de plantes hypométhylés à des plantes sauvages (rétrocroisement) (Lippman et al. 2003a, Teixeira et al. 2009). Ces études ont pu mettre en évidence de fortes disparités selon les éléments. En effet, alors que certains transposons réactivés dans le mutant hypométhylé restent actifs, d'autres sont à nouveau réprimés, partiellement ou totalement, dans les plantes F1, en association avec une reméthylation de l'élément. De plus, selon le type de méthylation affecté dans le parent mutant, les effets diffèrent. De façon intéressante, alors les mutants *ddm1* et *met1* dérèpriment tous deux tous les éléments analysés, la restauration de la fonction MET1 est capable de rééteindre partiellement une partie d'entre eux, tandis que dans les plantes F1 *ddm1*xWT aucun d'entre eux n'est réprimé (Lippman et al. 2003a, Teixeira et al. 2009) (**Figure I-29A**).

Parallèlement, une autre étude a démontré que certains transposons du knob du chromosome 4 étaient reméthylés et réprimés dans les lignées F5 issues d'autofécondations successives des plantes F1 *ddm1*xWT. Ces loci ont la particularité de conserver un taux non négligeable de méthylation ADN dans le mutant, ce qui permettrait d'amorcer la reméthylation impliquant de petits ARN de 24 nt (**Figure I-29B**) (Teixeira et al. 2009).



**Figure I-29 : Impact du rétrocroisement de plantes hypométhylées sur les profils de méthylation ADN.**

**A- Héritabilité de l'état activé de certains éléments transposables après rétrocroisement.** De nombreux éléments transposables activés dans les mutants *met1* ou *ddm1* demeurent fortement exprimés après rétrocroisement avec des plantes sauvages.

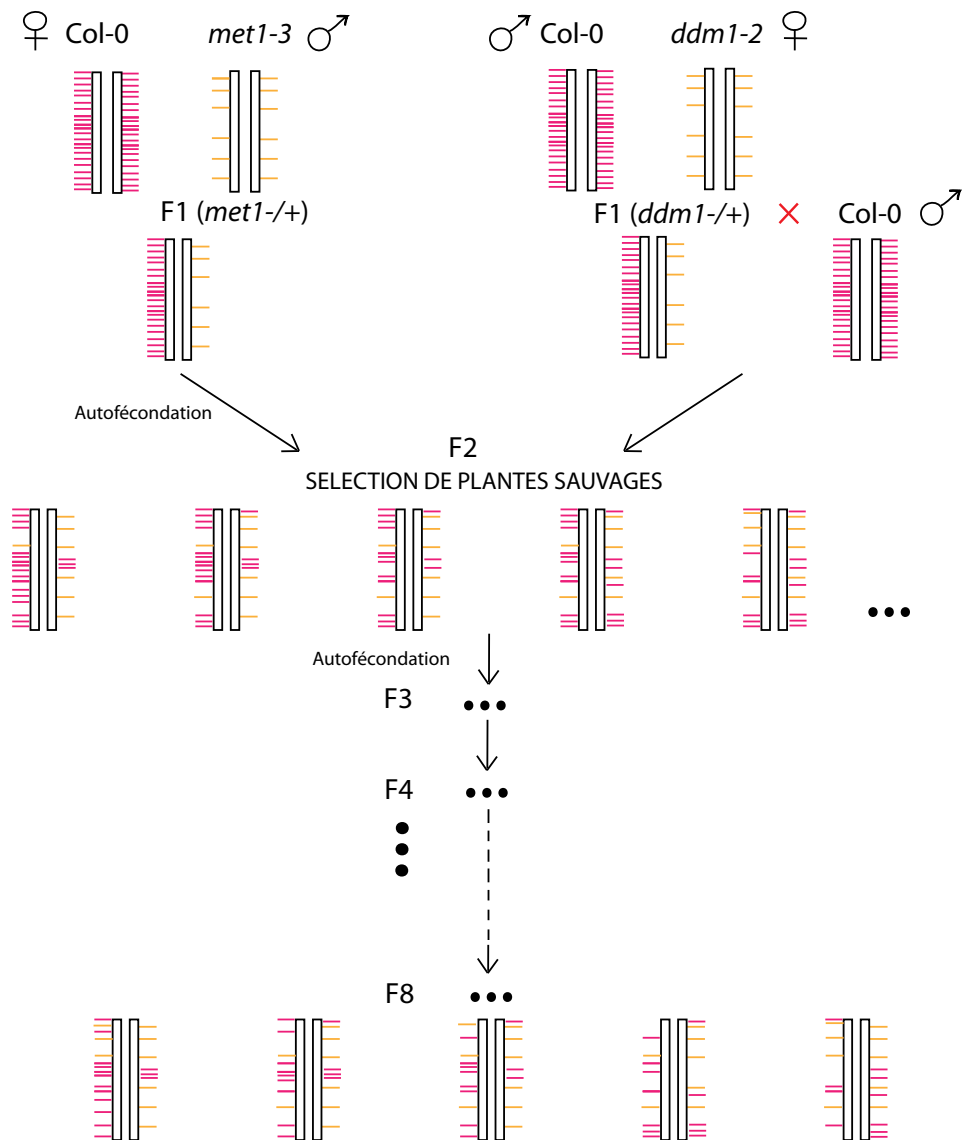
**B- Resilencing de certains éléments lors du rétrocroisement *ddm1* x WT.** Certains éléments transposables, activés dans *ddm1*, sont réprimés dans les générations suivant leur rétrocroisement. Ces éléments tendent à accumuler un fort taux de siRNA de 24 nt dans les plantes sauvages et à conserver un taux non négligeable de siRNA ainsi que de méthylation ADN dans le mutant *ddm1*. Ces siRNA pourraient permettre, via la voie RdDM, la reméthylation progressive et la répression de ces éléments.

Les siRNA pourraient également être impliqués dans la restauration du silencing de certains transposons dans les plantes F1 *met1*-/+, ainsi que l'H3K9me2, cette marque histone restant présente dans le mutant seulement sur les loci resilencés en F1 (Lippman et al. 2003a, Teixeira et al. 2009).

Les mutants *cmt3* et *kyp*, déficients en méthylation CHG ou en H3K9me2, ne réactivent, eux, que peu d'éléments, et la restauration de la fonction de ces protéines en F1 ne provoque pas la répression des transposons testés (Lippman et al. 2003a, Teixeira et al. 2009). Cependant, la complémentation de plantes mutantes *cmt3* par l'introduction d'un transgène *CMT3* permet de restaurer le silencing du gène *PAI2* issu du locus *PAI* répété et réarrangé ainsi que du gène *SUP* (Bartee et al. 2001, Lindroth et al. 2001). La présence de siRNA homologues générées à partir des répétitions inversées de ces deux loci, *PAI2* et *SUP*, pourrait participer au processus tandis qu'aucun siRNA n'a pu être détecté sur les éléments transposables réactivés dans *cmt3*.

### I.B.5. Analyse de lignées épigénétiques recombinantes

Afin d'étudier l'impact de la méthylation ADN sur la variabilité phénotypique, des lignées recombinantes épigénétiques (epiRILs -epigenetic recombinant inbred lines) ont été générées à partir d'hybrides F1 issus du rétrocroisement des mutants *met1* et *ddm1* (**Figure I-30**) (Johannes et al. 2009b, Reinders et al. 2009a). Les plantes sauvages obtenues en F2 (ou après un nouveau rétrocroisement dans le cas de *ddm1*) ont été sélectionnées et propagées sur six générations par autofécondation. Ainsi, a été obtenue une population de plantes qui sont isogéniques mais portant des profils épigénétiques distincts. Ces lignées ont montré une variation phénotypique importante, non seulement entre elles, mais également au cours des générations d'une lignée donnée. Des anomalies, telles qu'un retard de floraison, de croissance ou une résistance/sensibilité accrue aux stress biotiques et abiotiques, ont été notées dans certaines lignées. Le comportement des deux populations epiRILs diffère tout de même. En effet, de nombreux phénotypes instables apparaissent lors de l'autofécondation des epiRILs dérivés de *met1* (epiRILs-*met1*) (Johannes et al. 2009b, Reinders et al. 2009a). De plus, 30% de ces lignées seulement ont pu être propagées jusqu'en F8, contre 99% des lignées dérivées de *ddm1*.



**Figure I-30 : Mode de création des lignées épigénétiques isogéniques recombinantes (epiRILs).**

Les epiRILs sont issus du croisement de plantes sauvages avec des plantes hypométhylées (*ddm1* ou *met1*). Après autofécondation ou second rétrocroisement des plantes F1, les plantes sauvages sont sélectionnées parmi les individus F2 puis autofécondées jusqu'à obtention de lignées F8.

Certains caractères apparaissant dans les lignées d'epiRILs-*met1* sont ensuite hérités de façon stable. C'est pourquoi, l'héritabilité de la méthylation ADN des parents a été analysée dans des lignées F8 par hybridation sur puces d'échantillons d'ADN traités au bisulfite de sodium (Johannes et al. 2009b, Reinders et al. 2009a). Ceci a permis de mettre en évidence trois types de profils dans les epiRILs. Tout d'abord, des patrons de méthylation ADN intermédiaires ont été observés, indiquant un maintien inattendu d'épi-hétérozygotie au niveau de ces régions. De nombreux loci présentent à l'inverse des profils similaires au parent WT ou au parent *met1*, les épiallèles parentaux auraient donc être transmis de façon stable. Certaines régions chromosomiques dérivés de *met1* montrent un taux de méthylation semblable au sauvage, et ont du été reméthylés. De manière similaire, un mécanisme de reméthylation impliquant les petits ARN a été mis en évidence dans les lignées F5 *ddm1xWT* (Lippman et al. 2003a, Teixeira et al. 2009). Enfin, des épiallèles non-parentaux transgressifs, présentant un taux de méthylation supérieur ou inférieur aux deux parents ont également été détectés, ceci quelque soit l'origine parentale des segments chromosomiques affectés. Cette observation laisse supposer des variations des profils de méthylation lors de la rencontre initiale des épiallèles Col et *met1* et/ou durant les autofécondations successives. De plus, les nouveaux épiallèles ne sont pas toujours fixés et hérités de façon stable dans les générations suivantes. Ainsi, une très forte variation épigénétique est maintenue, ce qui pourrait expliquer l'instabilité phénotypique.

Parallèlement aux modifications profondes des profils épigénétiques dans les lignées epiRILs, de la variation génétique a été mise en évidence. En effet, certains éléments transposables, notamment des éléments de la famille CACTA sont mobilisés dans ces plantes épi-recombinantes. Ces transposons ADN sont transcrits et actifs dans le parent *ddm1* et continueraient donc à sauter dans les lignées epiRILs dérivés de ce mutant (Johannes et al. 2009b, Reinders et al. 2009a). La mobilisation de ces éléments dans les epiRILs-*met1* s'initie, elle, au cours de la génération de ces lignées puisque les transposons CACTA sont éteints et immobiles dans le parent sauvage comme dans le parent *met1* (Kato et al. 2004). La réactivation de ces éléments requérant une perte de méthylation CG et CHG (Kato et al. 2003b), ces données laissent penser que la mobilisation de ces éléments dans les lignées epiRILs-*met1* est causée par l'apparition d'épiallèles non-parentaux hypométhylés.





## I.C. Objectifs de la thèse

L'équipe d'Olivier Mathieu s'est particulièrement intéressée depuis plusieurs années au rôle de la méthylation CG et de MET1 dans la régulation des profils épigénétiques et transcriptionnels des gènes et éléments transposables. A travers l'analyse du mutant *met1*, il a pu participer à démontrer son impact sur les marques histone H3K4me, H3K9me ou H3K27me, sur le ciblage par la voie RdDM ou encore sur la déméthylation ADN, via la régulation des gènes *ROS1* et *DME*.

Mon travail de thèse s'est installé dans la continuité de ces recherches. En effet, dans le mutant *met1*, de très nombreux gènes présentent de la méthylation CHGme ectopique, ne s'expliquant ni par des anomalies de ciblage de la voie RdDM, ni par la dérégulation des déméthylases ADN précédemment rapportées. Une partie de mon étude a visé à déterminer la cause de cette hyperméthylation CHG. Nous avons noté que ce profil aberrant de méthylation était similaire à celui décrit dans le mutant *ibm1* et avons étudié le rôle de la méthylation ADN dans le contrôle du gène *IBM1* (Miura et al., 2009).

Les résultats obtenus durant ces recherches nous ont également permis de mettre en évidence une hypométhylation inattendue du gène *IBM1* au niveau des sites CHG lors du rétrocroisement des mutants *met1*. Le deuxième objectif de ma thèse était double, il visait donc à déterminer si ce phénomène affectait d'autres loci et, plus globalement, de décrire l'impact de la rencontre d'épigénomes distincts sur les profils de méthylation ADN et les profils transcriptionnels des plantes F1 *met1* x WT. En effet, la stabilité des profils de méthylation ADN au cours des générations avait été étudiée précédemment par la création de lignées épIRILs à partir des mutants *ddm1* et *met1* (voir paragraphe I.B.5). Nous souhaitions définir si les épiallèles non parentaux obtenus dans ces lignées résultaient d'un « choc épigénomique » ayant eu lieu lors du croisement initial entre ces mutants hypométhylés et les plantes sauvages, comme en avait émis l'hypothèse Eric Richards en 2009 (Richards 2009).

Ceci a été possible grâce à la réalisation d'un séquençage ARN, un séquençage bisulfite du génome entier ainsi qu'un séquençage des petits ARN dans Col, *met1* et les hybrides F1. Ainsi, nous avons obtenu une vue globale du paysage transcriptionnel et épigénétique dans ces épihybrides F1. Pour préciser l'ampleur de la restructuration du génome lors de la rencontre des épigénomes, nous avons également recherché des événements de transposition d'éléments transposables dans les plantes mutantes *met1* ainsi que dans la descendance des épihybrides.



## **II. Chapitre 1 : Etude de la régulation du gène *IBM1* dans le mutant *met1***

### **ARTICLE**

**“DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns”**

Rigal M\*, Kevei Z, Pélissier T, Mathieu O,  
The EMBO Journal, 2012



## II.A. Introduction

Cet article révèle le rôle de la méthylation ADN dans l'expression du gène *IBM1*, codant pour une déméthylase histone H3K9. Ma première année de thèse a été en grande partie dévolue au projet de recherche ayant abouti à la rédaction de cet article.

Ainsi, j'ai contribué aux analyses expérimentales, en collaboration avec Z. Kevei et T. Pélissier. J'ai notamment analysé les taux de H3K9me2 par Immunoprécipitation de Chromatine (ChIP), les taux de méthylation ADN par séquençage bisulfite, l'expression des loci par RT-PCR. J'ai aussi réalisé l'étude *in silico* de la structure des gènes orthologues d'*IBM1*.

En collaboration avec O. Mathieu et T. Pélissier, j'ai participé à la réflexion sur le projet ainsi qu'à la réalisation des figures et la rédaction du manuscrit, puis aux modifications suite aux commentaires des reviewers.



## II.B. Article « Rigal et al., 2012 »

The EMBO Journal (2012) 00, 1–13 | © 2012 European Molecular Biology Organization | All Rights Reserved 0261-4189/12  
www.embojournal.org

THE  
EMBO  
JOURNAL

## DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns

Mélanie Rigal, Zoltán Kevei,  
Thierry Pélissier and Olivier Mathieu\*

Centre National de la Recherche Scientifique (CNRS), UMR6293, GReD, INSERM U 1103, Clermont Université, Aubière, France

The stability of epigenetic patterns is critical for genome integrity and gene expression. This highly coordinated process involves interrelated positive and negative regulators that impact distinct epigenetic marks, including DNA methylation and dimethylation at histone H3 lysine 9 (H3K9me2). In *Arabidopsis*, mutations in the DNA methyltransferase MET1, which maintains CG methylation, result in aberrant patterns of other epigenetic marks, including ectopic non-CG methylation and the relocation of H3K9me2 from heterochromatin into gene-rich chromosome regions. Here, we show that the expression of the H3K9 demethylase IBM1 (increase in *BONSAI* methylation 1) requires DNA methylation. Surprisingly, the regulatory methylated region is contained in an unusually large intron that is conserved in *IBM1* orthologues. The re-establishment of *IBM1* expression in *met1* mutants restored the wild-type H3K9me2 nuclear patterns, non-CG DNA methylation and transcriptional patterns at selected loci, which included DNA demethylase genes. These results provide a mechanistic explanation for long-standing puzzling observations in *met1* mutants and reveal yet another layer of control in the interplay between DNA methylation and histone modification, which stabilizes DNA methylation patterns at genes.

The EMBO Journal advance online publication, 11 May 2012;  
doi:10.1038/emboj.2012.141

Subject Categories: chromatin & transcription

Keywords: DNA methylation; epigenetic stability; gene-body methylation; H3K9 demethylase; intron methylation

### Introduction

Silent chromatin is typically associated with specific patterns of epigenetic modifications, which in plants include high levels of DNA methylation in all three cytosine contexts (CG, CHG and CHH, where H = A, T or C) and dense dimethylation at lysine 9 of histone H3 (H3K9me2) (Bender, 2004; Chan *et al.*, 2005; Grewal and Jia, 2007). H3K9 methylation is conserved from plants to mammals and

relies on the activities of histone lysine methyltransferases in the Su(var)3-9 family (Rea *et al.*, 2000). The *Arabidopsis* genome encodes nine homologues of the *Drosophila* Su(var)3-9 protein, which are referred to as SUVH proteins (for Su(var)3-9 homologues) (Baumbusch *et al.*, 2001). Although SUVH4 (also known as KRYPTONITE or KYP), SUVH5 and SUVH6 seem to act redundantly in the maintenance of H3K9me2 at transposable elements (TEs) and repeats, only KYP appears to function at genes (Jackson *et al.*, 2004; Ebbs and Bender, 2006; Inagaki *et al.*, 2010). Conversely, enzymes of the JHDM2 family contain a jumonji C (jmc) domain and can remove H3K9 methylation (Klose *et al.*, 2006; Tsukada *et al.*, 2006; Yamane *et al.*, 2006). In *Arabidopsis*, experimental evidence supports the hypothesis that the jmc domain-containing protein increase in *BONSAI* methylation 1 (IBM1) is a histone demethylase that is specific for H3K9me2 and H3K9 monomethylation (Saze *et al.*, 2008; Inagaki *et al.*, 2010). Mutants for *IBM1* display ectopic H3K9me2 accumulation in the transcribed regions of a large number of genes, whereas TEs are unaffected (Inagaki *et al.*, 2010).

In *Arabidopsis*, CG methylation is propagated during DNA replication by the maintenance DNA methyltransferase METHYLTRANSFERASE 1 (MET1), which robustly copies methylation patterns on newly synthesized DNA strands. The maintenance of asymmetrical CHH methylation is mostly ensured by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) in a process known as RNA-directed DNA methylation (RdDM), which involves the polymerases IV and V (Law and Jacobsen, 2010). The perpetuation of CHG methylation patterns is largely ensured by the plant-specific chromomethylase CMT3, and genetic analyses suggest that targeting of CMT3 to chromatin relies on H3K9me2, which indicates that H3K9me2 acts upstream of CHG methylation (Lindroth *et al.*, 2004; Feng and Jacobsen, 2011). These two repressive marks are intimately associated, and at the genome level, ~90% of CHG methylation coincides with H3K9me2-enriched regions (Bernatavichute *et al.*, 2008). Additionally, the loss-of-function *kyp* and *cmt3* alleles show a similar loss of cytosine methylation at CHG sites and induce transcriptional reactivation of a common set of silent targets (Jackson *et al.*, 2002; Lippman *et al.*, 2003; Lindroth *et al.*, 2004; Ebbs *et al.*, 2005; Tran *et al.*, 2005; Ebbs and Bender, 2006). The SET and RING-associated (SRA) domain of KYP and SUVH6 binds to DNA that is methylated at CHGs *in vitro*, which suggests that CHG methylation also feeds back onto H3K9me2 (Johnson *et al.*, 2007).

In heterochromatin, methylation at CG sites and H3K9me2 are also linked. In wild-type (WT) plant nuclei, H3K9me2 is largely confined to heterochromatic chromosomal regions that are densely CG methylated. In the nuclei of the loss-of-function *met1-3* mutant, CG methylation is lost, and H3K9me2 is redistributed away from the chromocentres

\*Corresponding author. Centre National de la Recherche Scientifique (CNRS), UMR6293—INSERM U1103—GReD, Clermont Université, 24 Avenue des Landais, BP 80026 63171 Aubière Cedex, France.  
Tel.: +33 473 407 405; Fax: +33 473 407 777;  
E-mail: olivier.mathieu@univ-bpclermont.fr

Received: 14 October 2011; accepted: 20 April 2012





(Soppe *et al*, 2002; Tariq *et al*, 2003; Mathieu *et al*, 2007). Therefore, although the molecular mechanism remains elusive, CG methylation appears to direct H3K9me2 in heterochromatin. Importantly, CG methylation is not restricted to heterochromatin, and genome-wide methylation profiling studies have highlighted that approximately one-third of *Arabidopsis* genes are CG methylated (Tran *et al*, 2005; Zhang *et al*, 2006; Zilberman *et al*, 2007; Cokus *et al*, 2008; Lister *et al*, 2008). Noticeably, gene-body methylation is almost exclusively restricted to CG sites, and it is not associated with H3K9me2 (Bernatavichute *et al*, 2008). Therefore, CG methylation and H3K9me2 show distinct interactions at different chromosomal locations.

In addition, *met1* mutants exhibit ectopic non-CG methylation (Soppe *et al*, 2002; Tariq *et al*, 2003; Zhang *et al*, 2006; Mathieu *et al*, 2007; Cokus *et al*, 2008; Lister *et al*, 2008). At heterochromatic sequences, this has been shown to result primarily from the misdirection of the RdDM pathway, while aberrant methylation at a few hundred genes likely originates from the transcriptional downregulation of the DNA demethylases (which notably include *REPRESSOR OF SILENCING 1* (*ROS1*)), in the absence of CG methylation (Huettel *et al*, 2006; Mathieu *et al*, 2007). Importantly, several thousand genes specifically display CHG hypermethylation in their body sequence in the *met1* background, and the majority of these genes contain CG methylation in the WT background (Lister *et al*, 2008; Reinders *et al*, 2008). This suggests that CG methylation (and/or MET1 itself) may exclude CHG methylation from genes; however, the molecular mechanism that links CG and CHG methylation at genes remains to be elucidated. The recent analyses of the *IBM1* loss-of-function mutant have revealed widespread ectopic CHG DNA methylation and H3K9me2 at genes (Saze *et al*, 2008; Miura *et al*, 2009; Inagaki *et al*, 2010). These new epigenetic patterns are dependent on the function of CMT3 and KYP and interestingly, genes that contain CHG hypermethylation in the *met1* and *ibm1* mutant backgrounds largely overlap (Miura *et al*, 2009).

These similarities between the *met1* and *ibm1* mutants have led us to hypothesize that CG methylation and/or MET1 may protect genes from ectopic CHG methylation and H3K9me2 because they are required for proper *IBM1* expression. *IBM1* encodes two mRNA variants; the longer variant (*IBM1-L*) specifically encodes the functional IBM1 protein that contains the jmjC domain. We found that the proper accumulation of *IBM1-L* mRNA is controlled by DNA methylation and depends on the simultaneous presence of CG and CHG methylation in an unusually large intron of the *IBM1* gene. The re-establishment of *IBM1-L* expression in *met1* mutants largely suppressed the abnormal H3K9me2, DNA methylation and transcriptional patterns that were induced by the mutation at selected target genes. Interestingly, the expression of the *ROS1* DNA demethylase was recovered when *IBM1-L* accumulation was restored in the *met1* background, which thereby also contributed to the removal of aberrant DNA methylation patterns that occur in this mutant background. Therefore, by controlling the proper expression of H3K9 and DNA demethylases, CG methylation insures the maintenance of proper genic DNA methylation and histone modification patterns through a self-regulatory loop. These results highlight the importance of CG methylation as a

central coordinator of epigenetic stability at genes and provide mechanistic explanations for long-standing enigmatic observations in *met1* mutants.

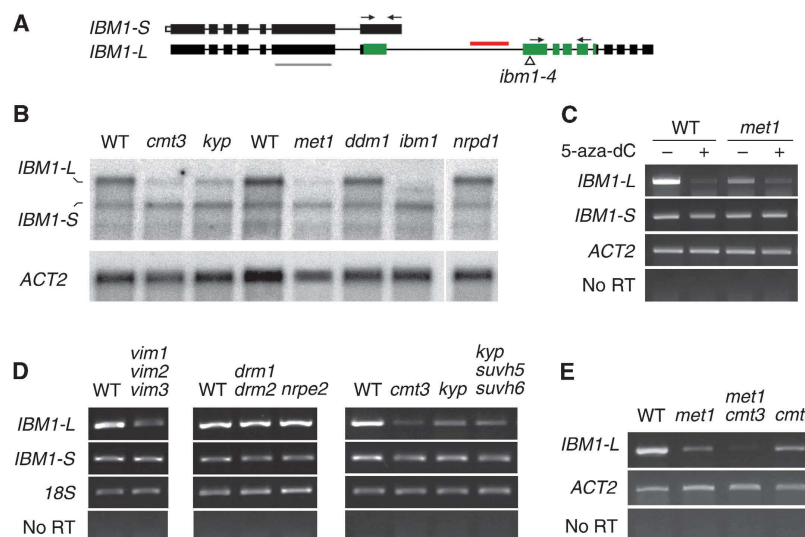
## Results

### DNA methylation is required for proper *IBM1* expression

To understand how CHG gene-body methylation occurs in *met1* mutant plants, we examined the transcription of the *IBM1* gene with respect to DNA methylation. *IBM1* is predicted to produce two different transcripts; only the longer transcript (hereafter referred to as *IBM1-L*) is predicted to encode the jmjC domain and thus the functional IBM1 protein (Figure 1A). Northern blotting and reverse transcription (RT)-PCR analyses confirmed the presence of the two predicted *IBM1* transcripts (Figure 1B and C). In WT plants that were treated with the DNA methylation inhibitor 5-aza-2'-deoxycytidine (5-aza-dC) and in plants that were homozygous for the *met1-3* null *MET1* allele (Saze *et al*, 2003), the accumulation of *IBM1-L* mRNA was specifically downregulated, whereas the accumulation of the short *IBM1* RNA transcript (*IBM1-S*) was not significantly affected (Figure 1B and C). To further investigate the impact of DNA methylation on *IBM1* expression, we assayed transcript accumulation in mutants of additional regulators of genomic DNA methylation. Triple mutants for the *VARIATION IN METHYLATION* (*VIM*) 1, 2 and 3 genes, which encode co-factors that are required for CG methylation maintenance (Woo *et al*, 2007, 2008; Kraft *et al*, 2008), also showed specific downregulation of the *IBM1-L* mRNA transcript (Figure 1D). Together with the results from the *met1* mutants, this suggests that CG methylation is required for the proper accumulation of *IBM1* transcripts that contain the putative H3K9 demethylase domain jmjC.

Noticeably, the downregulation of *IBM1-L* that was observed in the *met1* plants was further enhanced when the *met1* plants were grown on 5-aza-dC, which suggests that non-CG methylation may also participate in the control of *IBM1* expression (Figure 1C). In the *nrpd1* mutant of polymerase IV, as well as in the *nrpe2* (*nrpd2a*) mutant of the common subunit of polymerases IV and V, *IBM1* transcription was similar to the WT, and the *drm1 drm2* double mutant also exhibited no detectable *IBM1* transcriptional variation, which indicates that the RdDM pathway/CHH methylation does not control *IBM1* expression (Figure 1B and D). H3K9me2 and DNA methylation at CHG sites are intimately coupled, and a self-reinforcing feedback loop between KYP and CMT3 maintains these two marks in heterochromatin. We analysed the *cmt3* and *kyp* mutants to determine whether CHG methylation influences *IBM1* expression. RNA-gel blot and RT-PCR analyses showed that accumulation of the *IBM1-L* transcript was drastically downregulated in both *cmt3* and *kyp* (Figure 1B and D). Triple mutants of the histone H3K9 methyltransferases KYP(SUVH4)/SUVH5/SUVH6 mimic the *kyp* single mutant, indicating that *SUVH5* and *SUVH6* do not significantly contribute *IBM1* transcriptional control (Figure 1D). Together, these observations indicate that CG and CHG methylation are required for the proper expression of *IBM1*. This reveals another layer of interdependence and control between DNA methylation and H3K9 methylation, in which, somehow paradoxically with their role in the maintenance of DNA methylation, MET1, CMT3 and KYP are also





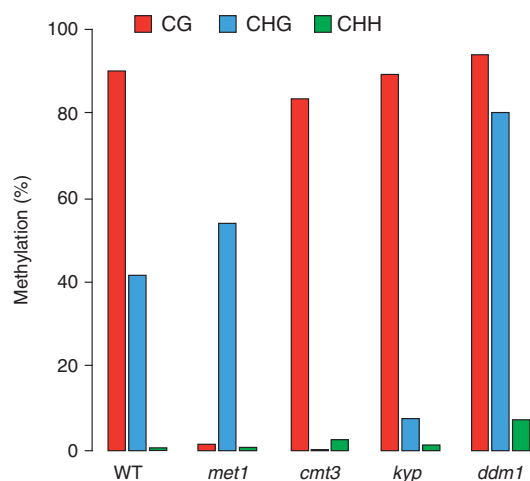
**Figure 1** DNA methylation is required for *IBM1-L* transcript accumulation. (A) Schematic representation of the two *IBM1* mRNA variants, *IBM1-S* and *IBM1-L*. The boxes represent exons; coding regions are indicated in black and untranslated regions are indicated in white. The region encoding the jmjC domain is shown in green (Saze et al, 2008). The intronic DNA-methylated zone (horizontal red bar), the positions of the probe that was used for the northern blot analysis (horizontal grey bar) and the primers that were used for the RT-PCR analysis of the *IBM1-L* and *IBM1-S* variants (arrows) are shown. The position of the T-DNA insertion in *ibm1-4* is indicated with a triangle. (B) Northern blot analysis of the *IBM1* transcripts using poly(A)<sup>+</sup> RNAs for the indicated mutant genotypes. Hybridization with a probe corresponding to *ACT2* (*ACT2*) is shown as a loading control. (C) RT-PCR analysis of *IBM1-S* and *IBM1-L* transcripts in WT and *met1-3* plants that were grown on medium containing (+) or lacking (-) 5-aza-dC. (D) RT-PCR analysis of the *IBM1-S* and *IBM1-L* transcripts in the indicated DNA methylation-deficient mutant backgrounds. (E) RT-PCR analysis of *IBM1-L* expression in *met1 cmt3* compared with the WT and single mutants. The amplification of *ACT2* or 18S rRNA (18S) was used to normalize the RNA template levels. The negative controls (no RT) lacked reverse transcriptase.

involved in the exclusion of CHG methylation/H3K9me2 from genes. The mutation of *DECREASE IN DNA METHYLATION 1* (*DDM1*) did not alter *IBM1* transcript accumulation (Figure 1B).

#### Intronic DNA methylation controls *IBM1* expression

To understand how DNA methylation controls *IBM1* expression, we examined DNA methylation profiles of the *IBM1* gene by bisulphite sequencing. It has been reported that gene-body methylation is preferentially targeted to nucleosomes on exons (Chodavarapu et al, 2010), and accordingly, *IBM1* carries body methylation at CG sites in several exons in WT plants (<http://neomorph.salk.edu/epigenome/epigenome.html>). In Arabidopsis, the average intron size is ~180 bp (The Arabidopsis Genome Initiative, 2000). The *IBM1* gene contains an unusually large intron that is >2 kb; this intron must be spliced out to generate the *IBM1-L* mRNA that encodes the jmjC histone demethylase domain. Interestingly, the large intron contains a zone that is densely methylated at both CG and CHG positions in the WT (<http://neomorph.salk.edu/epigenome/epigenome.html>; Figure 2). In the *met1-3* null mutant, exonic CG methylation was lost in the *IBM1* gene (<http://neomorph.salk.edu/epigenome/epigenome.html>). In the large intron, bisulphite sequencing confirmed that CG methylation was also erased; however, non-CG methylation, which was mostly represented by CHG methylation, was essentially maintained (Figure 2).

In the *kyp* mutant, CG methylation remained largely intact. However, methylation at CHG positions drastically decreased from ~40 to 10% (Figure 2). This indicates that CHG



**Figure 2** Bisulphite sequencing analysis of the DNA methylation pattern in the large intron of *IBM1*. The position of the methylated zone is indicated in Figure 1. The proportions of methylated cytosines at the CG (red), CHG (blue) and CHH (green) sites are given as percentages.

methylation in the large *IBM1* intron strongly depends on KYP activity. The remaining low level of CHG methylation in the mutant probably results from the redundant activity of other SUVH proteins, such as SUVH5 and/or SUVH6, which was previously shown for other loci (Ebbs et al, 2005; Ebbs



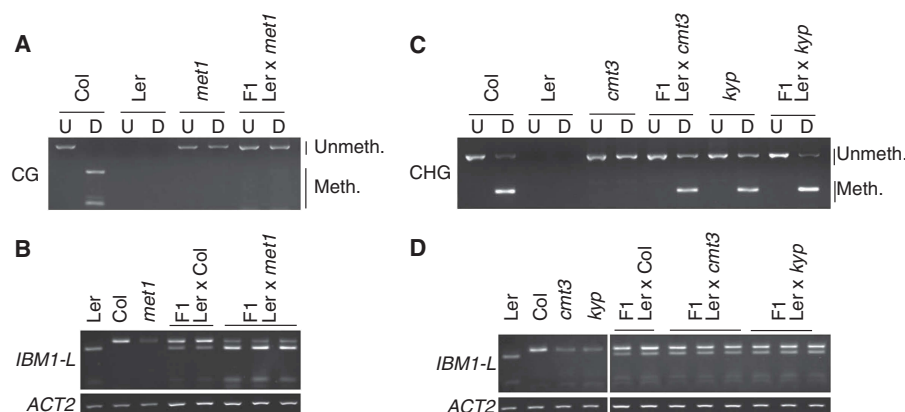
and Bender, 2006); although the contribution of SUVH5 and/or SUVH6 was not visualized at the transcriptional level (Figure 1D). DRM2 and CMT3 have been reported to be redundantly required for the maintenance of CHG methylation at particular genomic loci (Cao and Jacobsen, 2002; Cao *et al*, 2003). At the large intron of *IBM1*, CHG methylation was completely lost in *cmt3*, which indicates that CMT3 is solely responsible for CHG methylation at this locus; this is in agreement with the absence of detectable difference in *IBM1-L* accumulation in the *drm1 drm2* background (Figures 1D and 2).

Based on the impact of the *met1*, *cmt3* and *kyp* mutations on *IBM1-L* transcript accumulation, these data suggest that the simultaneous presence of CG and CHG methylation in the large *IBM1* intron is required for the proper production of *IBM1-L* transcripts. Accordingly, the *met1 cmt3* double mutant showed a stronger reduction in *IBM1-L* accumulation relative to each single mutant (Figure 1E). As mentioned above, the accumulation of *IBM1* transcripts was not altered by the *ddm1-2* mutation (Figure 1B), and consistent with our previous conclusion, the large *IBM1* intron was not hypomethylated but rather hypermethylated at all of the cytosine sequence contexts in this mutant (Figure 2). Regardless of the mechanism responsible for *IBM1* hypermethylation in the *ddm1* mutant background, this result supports the conclusion that DNA methylation in the large *IBM1* intron positively correlates with *IBM1-L* accumulation.

To further confirm that DNA methylation in the large *IBM1* intron controls the accumulation of RNA that encodes the protein that contains the H3K9me2 demethylase *jmjC* domain, we outcrossed the *met1*, *cmt3* and *kyp* mutants (all in the Col-0 background) with WT Ler-0 plants and assayed for DNA methylation and the transcription of the mutant-derived *IBM1* allele in the resultant F1 plants. Once they are altered in *met1* mutants, the CG methylation patterns cannot be re-established upon the reintroduction of MET1 activity (Soppe *et al*, 2000; Kankel *et al*, 2003; Saze *et al*, 2003; Mathieu *et al*, 2007). Consistently, PCR from bisulphite-treated DNA followed by digestion with a restriction enzyme showed

that the *met1*-derived *IBM1* allele was still unmethylated at CG sites in the Ler  $\times$  *met1* F1 hybrids (Figure 3A). This absence of remethylation correlated with a low accumulation of *IBM1-L* mRNA that originated from the *met1*-derived *IBM1* allele, which was similar to the *IBM1* transcription in the mutant parent (Figure 3B). In contrast to MET1, the reintroduction of CMT3 largely restores the developmental phenotypes induced by the loss of silencing that is associated with non-CG methylation in the *drm1/2 cmt3* mutants (Chan *et al*, 2006). Methylation at CHG sites, which was completely lost in the *cmt3* mutants, reappeared to a certain level in the large intron of the mutant-derived *IBM1* allele in the Ler  $\times$  *cmt3* F1 individuals (Figure 3C). RT-PCR analyses revealed that the accumulation of the *IBM1-L* transcript from the *cmt3*-derived allele was restored to WT levels, such that the *IBM1* transcriptional patterns in the Ler  $\times$  *cmt3* F1s and the Ler  $\times$  Col control F1s were indistinguishable (Figure 3D). Likewise, the accumulation of *IBM1-L* mRNA from the *kyp*-derived *IBM1* allele was re-established to WT levels in the F1 progeny of the Ler  $\times$  *kyp* cross (Figure 3D). Restoration of CHG methylation at the mutant-derived *IBM1* allele appeared to be more efficient in the Ler  $\times$  *kyp* F1s when compared with the Ler  $\times$  *cmt3* F1 plants (Figure 3C), and this was confirmed with bisulphite sequencing (Supplementary Figure S1). Because CHG methylation was completely lost in *cmt3* but not in *kyp*, CHG remethylation is likely more efficient or occurs faster in the presence of residual CHG methylation. Together, these observations indicate that DNA methylation in the large intron of *IBM1* is required for the proper accumulation of the *IBM1-L* transcript that encodes the *jmjC* domain.

RT-PCR analyses with primer sets that were designed along the length of the large *IBM1* intron revealed that the reduction of the *IBM1-L* transcript accumulation in the *met1* background occurred inside the methylated region of the large intron (Supplementary Figure S2). When primer sets located upstream of the intron methylation region were used, no clear difference in *IBM1* pre-mRNA level was detected between the WT and *met1*. Furthermore, northern blot



**Figure 3** Inheritance of *IBM1* intronic methylation and *IBM1-L* transcript accumulation. (A, C) DNA methylation of the *IBM1* large intron was assayed in the indicated genotypes by Col-0-specific PCR from bisulphite-treated DNA that was followed by digestion with *TaqI* (A) or *HpyCH4V* (C). CG and CHG methylation protect the *TaqI* and *HpyCH4V* sites, respectively, from bisulphite conversion, which facilitates restriction digestion after bisulphite treatment and PCR. Undigested (U) and digested (D) samples are shown. (B, D) Allele-specific RT-PCR analysis of *IBM1-L* transcript accumulation in Ler-0  $\times$  *met1* (B) Ler-0  $\times$  *cmt3* and Ler-0  $\times$  *kyp* (D) F1 individuals. The Ler  $\times$  Col-0 individuals are shown as controls. The *ACTIN 2* (*ACT2*) gene was used as a control.



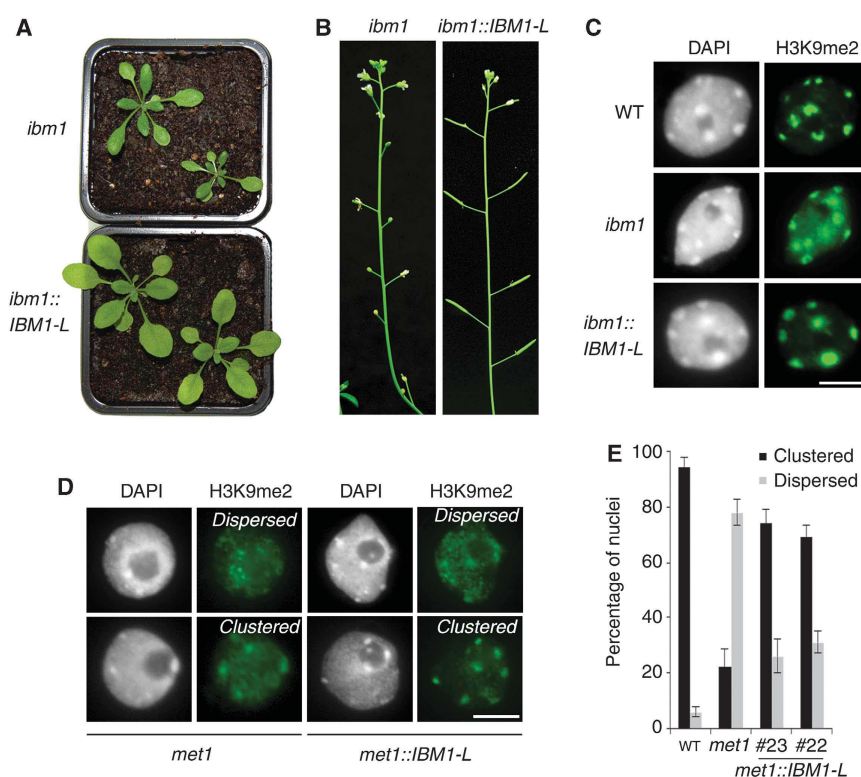
analyses using poly(A) + RNA samples revealed no stable alternative polyadenylated *IBM1* transcript variants in *met1*, *cmt3* or *kyp* relative to the WT, indicating no differential transcript polyadenylation in these mutants (Figure 1B). Collectively, these results favour the hypothesis that intronic DNA methylation at CG and CHG sites is required for proper *IBM1-L* transcript elongation.

**Enhanced *IBM1-L* mRNA accumulation complements the *ibm1* mutation and largely restores H3K9me2 patterns in *met1***

The loss of CG methylation in *met1* mutants results in aberrant patterns of other epigenetic marks, including non-CG methylation and H3K9me2. In *met1-3* mutant nuclei, H3K9me2 is relocated away from heterochromatic chromocentres into euchromatic chromosomal regions (Tariq *et al*, 2003; Mathieu *et al*, 2007). To date, the underlying molecular mechanism that is responsible for this relocation has not been elucidated. Because *IBM1* targets genes for H3K9 demethylation, we hypothesized that the lower accumulation of *IBM1-L* may account for the ectopic accumulation of H3K9me2 at genes in *met1*.

To test this hypothesis, we cloned the *IBM1-L* cDNA under the endogenous *IBM1* promoter and used the resulting

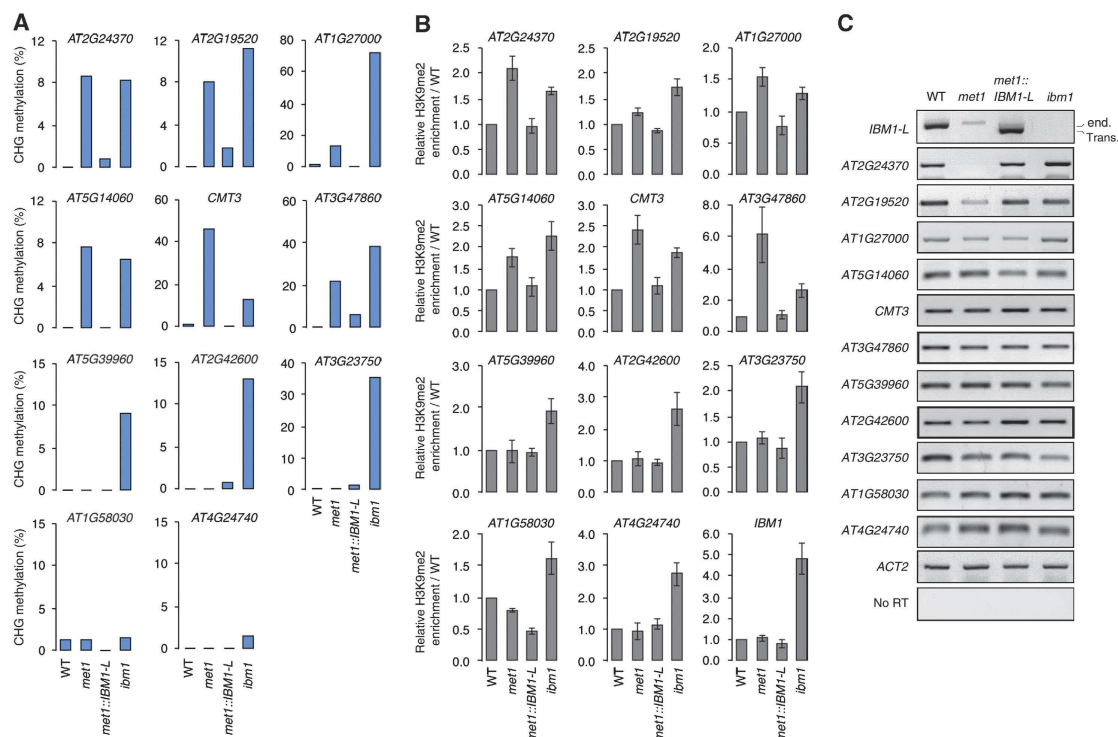
construct (*pIBM1:IBM1-L*) to transform *ibm1-4/IBM1* and *met1-3/MET1* heterozygous plants. Unlike untransformed *ibm1* mutant segregants, the *ibm1* individuals expressing *IBM1-L* (*ibm1::IBM1-L*) exhibited WT-like leaf size and morphology, as well as restored fertility (Figure 4A and B). Additionally, *ibm1::IBM1-L* plants showed reduced ectopic DNA methylation at the *BONSAI* gene (Supplementary Figure S3). These observations indicate that the *pIBM1:IBM1-L* construct efficiently complements the *ibm1-4* mutation. In WT nuclei, H3K9me2 is essentially clustered at heterochromatic chromocentres (Figure 4C and E). The H3K9me2 signal in euchromatic regions was notably enhanced in nuclei of the *ibm1-4* mutant compared with the WT (Figure 4C), which is consistent with the fact that *IBM1* targets a large number of genes for H3K9 demethylation (Inagaki *et al*, 2010). The nuclei of complemented *ibm1::IBM1-L* lines displayed H3K9me2 patterns that were indistinguishable from the WT (Figure 4C). In nuclei from first-generation *met1-3* mutants, H3K9me2 was frequently associated with euchromatic nuclear regions (Figure 4D and E; Tariq *et al*, 2003; Mathieu *et al*, 2007). The restored accumulation of *IBM1-L* in the *met1-3* background in *met1::IBM1-L* plants (see Figure 5C and Supplementary Figure S5) drastically increased the proportion of nuclei that exhibited WT-like H3K9me2 patterns; this



**Figure 4** *IBM1-L* expression complements the *ibm1-4* mutation and restores H3K9me2 patterns in *met1::IBM1-L* plants. (A) Representative images of 3-week-old siblings segregating from self-pollination of an *ibm1-4/IBM1::IBM1-L/+* parent are shown. (B) Inflorescences of *ibm1* and *ibm1::IBM1-L* plants. H3K9me2 patterns in WT, *ibm1-4*, *ibm1::IBM1-L* (C), *met1* and *met1::IBM1-L* plants (D) were analysed by immunocytology with a specific antibody against this mark. Representative images from one experiment are shown; two independent *ibm1::IBM1-L* and *met1::IBM1-L* (#22 and #23) lines were analysed, and they provided similar results. Two different H3K9me2 patterns were commonly observed in the *met1-3* and *met1::IBM1-L* plants, and the signals were either dispersed from or clustered at the chromocentres. (E) Proportion of both types of nuclei was monitored and is represented as a histogram ( $\pm$  s.d.). Between 60 and 80 nuclei from three independent experiments were scored. Scale bar = 5  $\mu$ m.







**Figure 5** Restoration of *IBM1-L* transcript accumulation in *met1* plants suppresses ectopic CHG hypermethylation and H3K9me2 enrichment at genes. (A) CHG methylation was determined in the WT, *met1-3*, *met1::IBM1-L* (#23) and *ibm1-4* at the body of the indicated genes using bisulphite sequencing. Each region analysed contains 11–23 CHG sites. The full DNA methylation profiles are provided in Supplementary Figure S4. (B) Association of the indicated genes with H3K9me2 was determined by ChIP with a specific antibody against this mark and is represented as a relative enrichment over the WT. Quantifications are from three independent experiments ( $\pm$  standard error of the mean). (C) RT-PCR analysis of transcript accumulation at the indicated genes. 'end.' endogenous *IBM1* transcript, 'trans.' transcript from the *IBM1-L* transgene. Amplification of *ACTIN2* (*ACT2*) was used to normalize the RNA template levels. The negative controls (no RT) lacked reverse transcriptase.

mark was preferentially associated with heterochromatin and less-intense euchromatin staining (Figure 4D and E). These results indicate that the downregulation of *IBM1-L* largely accounts for the relocation of H3K9me2 at gene-rich euchromatic regions that occurs in *met1* mutants.

#### Downregulation of *IBM1-L* is responsible for ectopic CHG methylation at genes in *met1* mutants

Because the maintenance of H3K9me2 and CHG methylation are mechanistically linked, we sought to determine whether the relocation of H3K9me2 in heterochromatin in *met1::IBM1-L* was accompanied by the suppression of ectopic CHG methylation at genes. Using bisulphite sequencing, we analysed DNA methylation at 11 genes that are body methylated at CG positions in the WT (<http://neomorph.salk.edu/epigenome/epigenome.html>). Methylation at CHG sites was virtually absent in the WT at all these genes (Figure 5A). In agreement with the fact that *IBM1* targets a large number of genes for H3K9me2 demethylation, all genes but one (*AT1G58030*) showed ectopic CHG methylation in the *ibm1* mutant background compared with the WT. Among these, six also exhibited ectopic CHG methylation ( $\sim 10$ –40%) in the *met1* background (Figure 5A; Supplementary Figure S4). In *met1::IBM1-L*, CHG methylation levels were strongly decreased in the body of these six genes, indicating that the

enhanced expression of *IBM1-L* in *met1* largely suppresses ectopic body methylation at CHG sites (Figure 5A; Supplementary Figure S4). Interestingly, four genes (*AT5G39960*, *AT2G42600*, *AT3G23750*, *AT4G24740*) were CHG hypermethylated in *ibm1* but not in *met1* (Figure 5A), suggesting that some genes are protected from ectopic CHG methylation in *met1*, although they are targeted for H3K9me2 demethylation by *IBM1*. Chromatin immunoprecipitation (ChIP) assays confirmed that the gene-body CHG hypermethylation in both *met1* and *ibm1* mutants was associated with enrichment in H3K9me2 at the same regions analysed for DNA methylation relative to the WT (Figure 5B). Consistent with the fact that *IBM1-L* encodes the functional form of the H3K9 demethylase, nearly WT H3K9me2 levels were restored in *met1::IBM1-L* at genes, showing ectopic H3K9me2 in the *met1* mutant background (Figure 5B). Interestingly, the intronic DNA-methylated region of the *IBM1* gene also showed H3K9me2 enrichment in *ibm1*, indicating that *IBM1* targets its large intron for H3K9me2 demethylation (Figure 5B). However, the persistence of high CHG methylation levels at this genomic region in WT plants (see Figure 2), suggests that KYP-mediated H3K9me2 deposition surpasses *IBM1*-mediated demethylation activity.

We next determined whether these changes in DNA methylation and H3K9me2 patterns were associated with changes

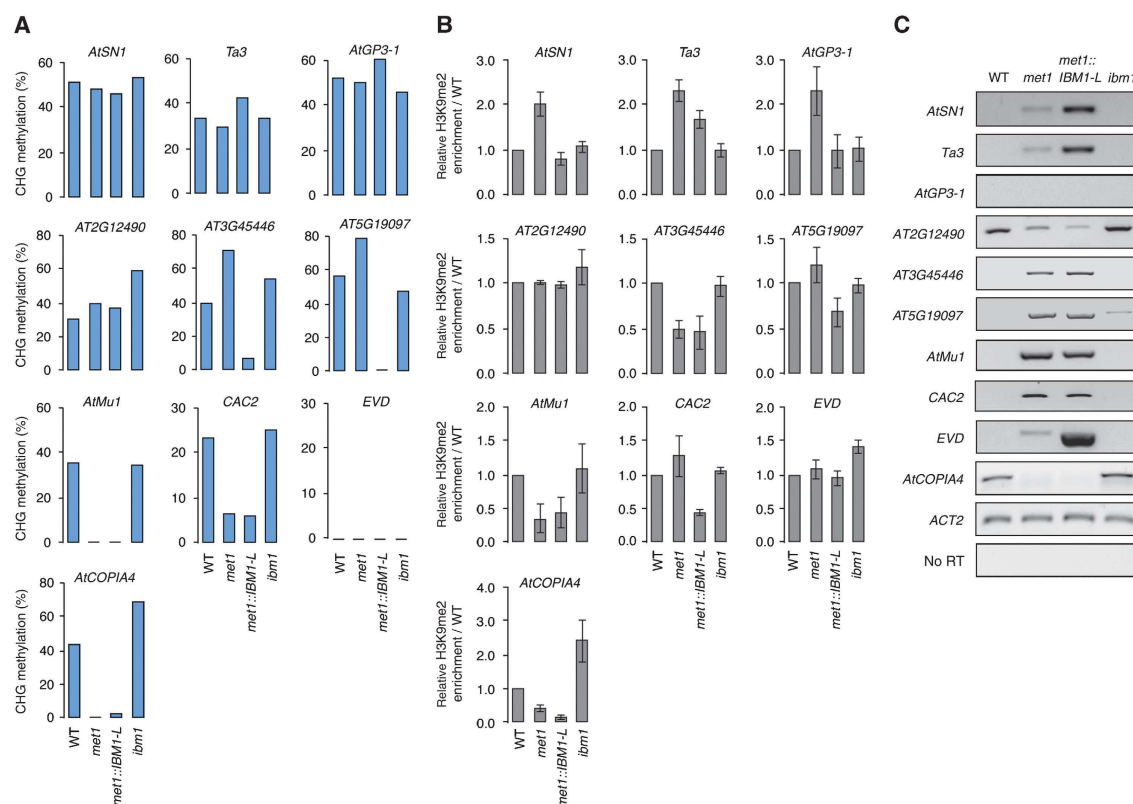


in transcriptional patterns. All 11 genes were transcribed in the WT and their transcription was not significantly altered in *ibm1*, although most of these genes showed ectopic CHG methylation and H3K9me2 in this mutant background and intact CG methylation (Figure 5A–C; Supplementary Figure S4). This suggests that aberrant CHG body methylation does not apparently impede transcription in the presence of WT CG gene-body methylation. Previous studies have revealed a modest impact of *met1* mutation on gene expression compared with the massive transcriptional activation of pseudo-genes and TEs (Zhang *et al*, 2006; Lister *et al*, 2008). Accordingly, of the six genes analysed showing ectopic CHG methylation in *met1*, only two (*AT2G24370* and *AT2G19520*) were transcriptionally downregulated in *met1* (Figure 5C; Supplementary Figure S5), which indicates that ectopic CHG body methylation does not necessarily lead to transcriptional silencing, even in the absence of CG methylation. Transcription from the genes that did not show ectopic CHG methylation in *met1* was also not detectably altered in this mutant background (Figure 5C).

Importantly, suppression of both ectopic CHG methylation and CG methylation at *AT2G24370* and *AT2G19520* in *met1::IBM1-L* (Figure 5A; Supplementary Figure S4) restored WT transcript accumulation from these two genes (Figure 5C; Supplementary Figure S5). Therefore, CG body methylation appears to act as a protective mark at certain genes, safeguarding their transcription from the deleterious

consequences of aberrant non-CG methylation. Why only a subset of genes do need such protection remains to be elucidated.

Unlike a large number of genes, it has been reported that TEs are not targeted by IBM1 (Inagaki *et al*, 2010). Accordingly, of the 10 elements that we analysed, only *AtCOPIA4* was enriched in H3K9me2 and CHG hyper-methylated in the *ibm1* mutant (Figure 6A and B). Methylation at CG sites was shown to direct H3K9me2 at heterochromatic sequences, and H3K9me2 depletion was observed at selected TEs in *met1* (Soppe *et al*, 2002; Lippman *et al*, 2003; Tariq *et al*, 2003). Although *AtMu1*, *AtCOPIA4*, *AT3G45446* showed markedly decreased H3K9me2 levels in *met1*, no/subtle changes (*AT5G19097*, *AT2G12490*, *CAC2*, *EVD*) or clear enrichment in H3K9me2 (*AtSN1*, *Ta3*, *AtGP3-1*) were observed at the other elements (Figure 6B), indicating that the dependence of H3K9me2 over CG methylation is not conserved between distinct TEs. Noticeably, H3K9me2 levels in *met1* and in *met1::IBM1-L* were highly similar, except at TEs showing increased H3K9me2 levels in *met1* (e.g., *AtGP3-1*; Figure 6B). At these elements, enhanced expression of *IBM1-L* in *met1::IBM1-L* resulted in decreased H3K9me2 compared with *met1* (Figure 6B). This suggests that these elements can be targeted by IBM1 for H3K9me2 demethylation in the absence of MET1/CG methylation. In contrast to genes, the changes in H3K9me2 levels at TEs were not necessarily mirrored by



**Figure 6** The mechanisms governing H3K9me2 and CHG methylation largely differ at genes and TEs. CHG methylation (A), H3K9me2 enrichment (B) and transcript accumulation (C) was determined at the indicated TEs in the WT, *met1-3*, *met1::IBM1-L* (#23) and *ibm1-4* as described in the legend of Figure 5. Quantifications shown in (B) are from three independent experiments ( $\pm$  standard error of the mean).



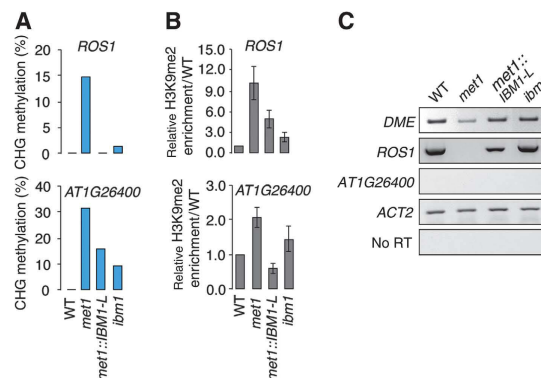
changes in CHG methylation. For instance, although *AtSN1*, *Ta3* and *AtGP3-1* were enriched in H3K9me2 in *met1*, CHG methylation remained nearly unchanged (Figure 6A).

RT-PCR analyses revealed that most of the elements were silent in the WT and transcriptionally activated in *met1* mutants except the *AtGP3-1* retrotransposon, which was still transcriptionally repressed in this mutant background as previously reported (Tsukahara *et al*, 2009; Figure 6C). Interestingly, silencing of *AtGP3-1* in *met1* suggests that IBM1 targeting at this locus occurs independent of transcription, although we cannot rule out transcriptional activity from this TE at specific developmental stages or in particular tissues of *met1* plants. Conversely, the active transcription from *AT2G12490* and *AtCOPIA4* that was detected in the WT was downregulated in the *met1* mutant background (Figure 6C), indicating that CG methylation differently impacts the transcription of distinct elements. These TEs were also transcribed in *ibm1*, suggesting that its transcriptional activity in the WT does not rely on IBM1 H3K9me2 demethylase activity. When compared with *met1*, the transcription of TEs was not significantly affected in *met1::IBM1-L*, except that of *AtSN1*, *Ta3* and *EVD*, which was upregulated in *met1::IBM1-L* (Figure 6C). However, since this transcriptional upregulation was not necessarily associated with H3K9me2 depletion or decreased CHG methylation at these elements, it probably results from indirect effects of IBM1 on loci restraining the transcription of these elements in the *met1* background. Together, these results confirm that IBM1 does not generally target TEs and highlight that the mechanisms governing H3K9me2 and CHG methylation largely differ at genes and TEs.

#### IBM1-L downregulation influences the expression of DNA demethylases

The Arabidopsis genome contains four genes that encode DNA demethylases, which are named *DEMETER* (*DME*), *ROS1/DEMETER-LIKE1* (*ROS1/DML1*), *DML2* and *DML3*. We previously showed that the expression of *ROS1* and *DME* is repressed at the transcriptional level upon the loss of CG methylation in the *met1-3* background (Mathieu *et al*, 2007). Because these demethylases preferentially target genic regions for demethylation (Penterman *et al*, 2007), their downregulation likely contributes to the aberrant DNA methylation patterns that are generated in *met1*. The effect of *DME* downregulation was confirmed by the sporadic appearance of abnormal flower phenotypes in *met1* lines; these phenotypes were highly reminiscent of the developmental abnormalities that were observed in some *met1 dme* homozygotes (Supplementary Figure S6; Xiao *et al*, 2003).

*DME* primarily functions in the central cell of the female gametophyte, whereas *ROS1* additionally acts in the sporophyte. Interestingly, similar to the other genes that were analysed above (Figure 5A), bisulphite sequencing revealed that the coding region of *ROS1* was ectopically hypermethylated at non-CG sites in *met1* (Figure 7A; Supplementary Figure S4). Moreover, this methylation correlated with H3K9me2 enrichment in *met1* and was directly associated with the downregulation of *IBM1-L* because it was suppressed in the *met1::IBM1-L* background (Figure 7B and C). The *DME* gene is similarly hypermethylated at CHG sites in *met1* plants (<http://neomorph.salk.edu/epigenome/epigenome.html>).



**Figure 7** IBM1 targets DNA demethylase genes in the *met1* background. (A) CHG methylation in the gene body of *ROS1* and a *ROS1*-target gene (*AT1G26400*) was determined in the WT, *met1-3* and a *met1::IBM1-L* line (#23) using bisulphite sequencing. The full DNA methylation profiles are provided in Supplementary Figure S4. (B) Association of *ROS1* and *AT1G26400* with H3K9me2 was determined by ChIP and is represented as a relative enrichment over the WT. Quantifications are from three independent experiments ( $\pm$  standard error of the mean). (C) RT-PCR analysis of the *ROS1*, *DME* and *AT1G26400* transcript accumulation in the indicated genotypes. The amplification of *ACTIN 2* (*ACT2*) was used to normalize the RNA template levels. The negative controls (no RT) lacked reverse transcriptase.

Importantly, the restored expression of *IBM1-L* in *met1* also restored *DME* transcription and resumed *ROS1* transcription, which was completely absent in the *met1-3* background (Figure 7C; Supplementary Figure S5). This result indicates that the downregulation of *IBM1-L* in *met1* mutants contributes to the creation of aberrant DNA methylation patterns at genes; this occurs because of its lack of direct activity at genes and also through the negative control of DNA demethylases expression. To further support this conclusion, we analysed DNA methylation at a target gene of *ROS1* (*AT1G26400*), which is hypermethylated in the *ros1-3* mutant (Penterman *et al*, 2007). As anticipated, this locus was also enriched in H3K9me2 and hypermethylated in *met1*, primarily at non-CG sites, but it was much less DNA hypermethylated and showed reduced H3K9me2 levels in *met1::IBM1-L* (Figure 7A; Supplementary Figure S4). This locus was enriched in H3K9me2 and CHG hypermethylated in the *ibm1* mutant (Figure 7A and B), indicating that it is also targeted for H3K9me2 demethylation by IBM1.

## Discussion

### Controlling gene expression through intron DNA methylation

DNA methylation has been widely associated with the repression of gene expression, for which two mechanisms have been proposed (Bird and Wolffe, 1999; Bird, 2002). The methylation of promoter sequences can directly inhibit transcription by blocking transcription factor binding; alternatively, DNA methylation can indirectly prevent transcription by recruiting various transcriptional repressors to form an inactive chromatin conformation. Here, we reveal an unorthodox role for DNA methylation, in which its presence in an intron is required for the proper expression of a gene. The large intron of *IBM1* contains a region that is



densely methylated at both CG and CHG positions, and a reduction in the level of either type of methylation results in the decreased accumulation of the *IBM1-L* transcript. In *Arabidopsis*, the average intron size is 180 bp (The *Arabidopsis* Genome Initiative, 2000). The *IBM1* gene contains an unusually large intron of over 2 kb that must be spliced out to generate the *IBM1-L* mRNA, which encodes the jmjC histone demethylase domain. Less than 1% of *Arabidopsis* introns are >1 kb (Wang and Brendel, 2006). Because large introns may require more energy for transcription and splicing, they are expected to be selected against during evolution unless they provide regulatory features (Castillo-Davis *et al*, 2002). *IBM1* orthologues in distantly related plant species and honeybee also contain introns that are much larger than the mean intron size of their respective genomes, which supports the evolutionary ancient origin of the large intron size in the regulation of *IBM1* expression (Supplementary Figure S7).

At least three hypotheses can be formulated to explain how intronic methylation may promote *IBM1-L* accumulation. First, DNA methylation may prevent premature transcription termination by obscuring transcription termination sites. Indeed, DNA methylation tends to be distributed away from the 3' ends of genes in *Arabidopsis*, suggesting that methylation is deleterious for transcription termination (Tran *et al*, 2005; Zhang *et al*, 2006; Zilberman *et al*, 2007; Lister *et al*, 2008). Second, DNA methylation may be required to maintain chromatin in an open state that is permissive for transcription elongation over the large intron. Specific methyl-binding proteins may bind to the methylated intronic region and recruit proteins that can modulate the chromatin compaction state, such as histone acetyltransferases. For example, the *Arabidopsis* METHYL-CpG BINDING DOMAIN 9 (AtMBD9) protein has been shown to bind to the *FLOWERING LOCUS C* (*FLC*) locus, where it triggers local histone acetylation, and enhances *FLC* expression (Peng *et al*, 2006; Yaish *et al*, 2009). A third possibility is that the presence of DNA methylation may directly or indirectly prevent the recruitment of a transcriptional repressor. In support of the latter hypothesis, DNA methylation was shown in mammals to interfere with the binding of the CCCTC-binding factor (CTCF), which induces the transcriptional repression of its target genes (Lobanenkov *et al*, 1990; Bell and Felsenfeld, 2000; Hark *et al*, 2000; Lai *et al*, 2010; Rodriguez *et al*, 2010). Moreover, CTCF binding influences polymerase II elongation dynamics and promotes RNA polymerase II pausing (Shukla *et al*, 2011). Although no direct CTCF homologues have been described in *Arabidopsis* (Heger *et al*, 2009), the loss of DNA methylation in the large *IBM1* intron may recruit proteins with similar repressor activities, which could induce polymerase II pausing or reduce its elongation rate in the large intron; this would thereby lead to the premature dissociation of the polymerase, which has already been demonstrated in yeast (Uptain *et al*, 1997; Mason and Struhl, 2005; Gromak *et al*, 2006), and would therefore result in low *IBM1-L* transcript accumulation. Although our results tend to suggest that intronic DNA methylation may indeed be required for proper *IBM1-L* transcript elongation, the identification of additional mutations that affect *IBM1* transcription will help to discriminate between these alternatives. Our results set the stage for future work to

determine the molecular mechanism by which intronic DNA methylation affects *Arabidopsis* gene expression. Interestingly, this mechanism may be conserved in other eukaryotes because the expression of the *EGR2* gene in mammals also appears to be positively regulated by intronic methylation (Unoki and Nakamura, 2003).

### CG methylation and epigenome stability

From observations in *Neurospora crassa* and *Arabidopsis*, it has been long understood that DNA methylation and H3K9 methylation interact in heterochromatin. In *N. crassa*, DNA methylation by the DIM-2 DNA methyltransferase depends on the methylation of H3K9, which is deposited by DIM-5 (Rountree and Selker, 2010). In *Arabidopsis*, the efficient maintenance of CHG methylation involves a self-reinforcing feedback loop between CMT3, which is structurally similar to DIM-2, and KYP (Johnson *et al*, 2007). Dense CG methylation, which is characteristic of heterochromatin, also directs H3K9me2 and the loss of CG methylation in *met1* null mutants leads to the depletion of H3K9me2 at heterochromatic sequences and the enrichment of H3K9me2 at euchromatic regions (Soppe *et al*, 2002; Tariq *et al*, 2003; Mathieu *et al*, 2007). Moreover, we and others previously showed that CG methylation directly prevents the accumulation of H3K4me2 and H3K27me3 in heterochromatin (Tariq *et al*, 2003; Mathieu *et al*, 2005; Zhang *et al*, 2009). In addition to the post-translational modification of histones, CG methylation also influences the deposition of histone variants, and there is a strong anti-correlative relationship between DNA methylation and H2A.Z distribution (Zilberman *et al*, 2008). Finally, CG methylation is required to stabilize proper patterns of non-CG methylation. In a manner that resembles CpG island methylation that occurs in mammalian tumorigenesis, the loss of CG methylation induces the CHH hypermethylation of several heterochromatic sequences and ectopic CHG methylation in the body of a large number of genes (Mathieu *et al*, 2007; Cokus *et al*, 2008; Lister *et al*, 2008; Reinders *et al*, 2008). The hypermethylation of heterochromatin was linked to the misdirection of the RdDM pathway, and the hypermethylation of several genes was attributed to inefficient DNA demethylation (Mathieu *et al*, 2007). Somehow paradoxically, the efficient transcription of genes that encode DNA demethylases, including *ROS1* and *DME*, requires CG methylation, such that *ROS1* transcription is totally impaired in *met1* null mutants (Mathieu *et al*, 2007).

Mechanistically, it is unclear how CG methylation affects all of these epigenetic factors and epigenetic marks. Here, we demonstrated that the proper expression of a gene that encodes another demethylase, the *IBM1* H3K9 demethylase, surprisingly also requires DNA methylation. Through this effect, CG methylation directly contributes to the stability of DNA methylation patterns at genes. In the absence of *MET1*, the loss of CG methylation in the large intron of *IBM1* results in the downregulation of the functional form *IBM1-L*, which demethylates a large number of genes. By restoring *IBM1-L* accumulation in the *met1-3* background, we showed that *IBM1-L* downregulation is responsible for the relocation of H3K9me2 in euchromatin and accounts for ectopic DNA methylation and H3K9me2 at genes in *met1* mutants. Ectopic genic DNA methylation results from at least two

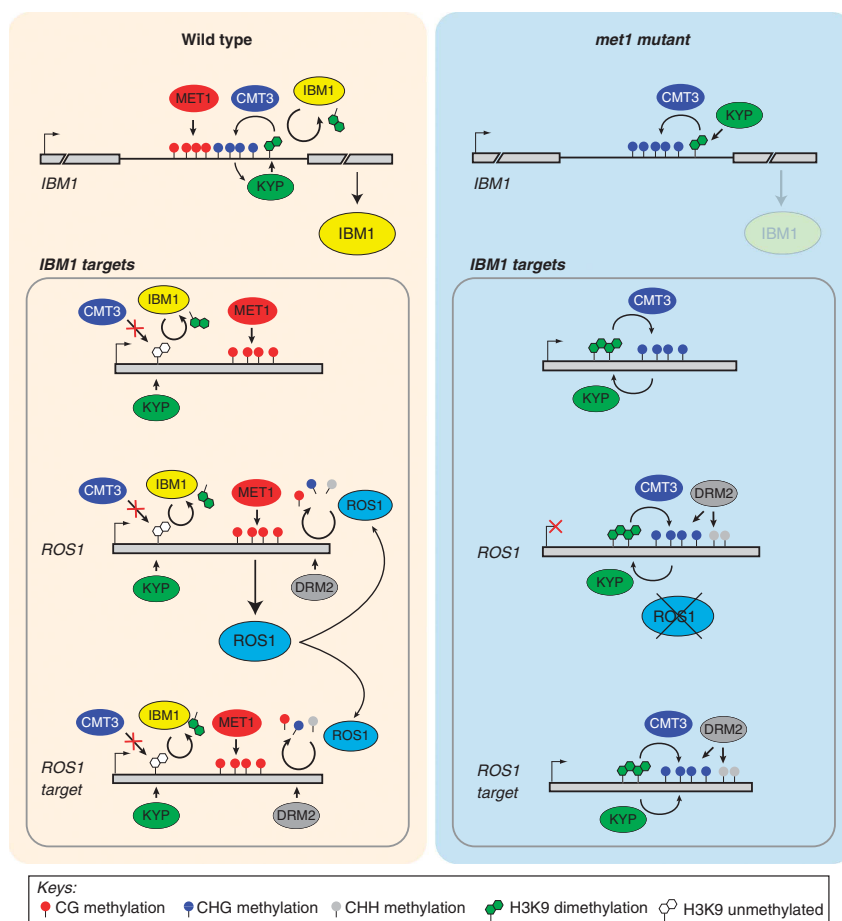




distinct causes at different loci: (i) decreased IBM1-mediated H3K9 demethylation and (ii) the loss of ROS1 DNA demethylase activity. Remarkably, we demonstrated that these two mechanisms are also interconnected because the restored *IBM1-L* expression in *met1* resumes *ROS1* transcription (Figure 8). Moreover, we showed that the functions of KYP and CMT3 are also required to promote proper *IBM1-L* accumulation. These interconnections illustrate the sophisticated interplay between the methylation and demethylation pathways at the DNA and the histone levels.

Transcribed genes tend to be more CG methylated than poorly transcribed genes, and in *met1*, which displays reduced IBM1 activity, CG gene-body methylation is largely replaced by CHG methylation (Zhang *et al*, 2006; Zilberman *et al*, 2007; Lister *et al*, 2008). A comparison of CG body-

methylated genes in the WT (Lister *et al*, 2008) with genes that show ectopic DNA hypermethylation in *ibm1* mutants (Miura *et al*, 2009) revealed that the majority of genes most hypermethylated in *ibm1* (about 73 and 57% of Class I and Class II genes, respectively) are strongly enriched in methylated CG positions (>25 methylated CGs) compared with the proportion of all genes with such CG methylation density (about 23%; Supplementary Figure S8). Therefore, we propose that CG methylation is involved in the recruitment of IBM1 to genes in WT plants, and CHG methylation can substitute for this signal in the *met1* background. In support of this hypothesis, the CG and CHG methylation profiles are very similar in WT and *met1* plants, respectively (Lister *et al*, 2008). If DNA methylation is the signal, then why is IBM1 not recruited to silent, heavily DNA-methylated



**Figure 8** Model for the appearance of ectopic non-CG methylation and H3K9me2 at genes in *met1* mutants. In WT plants, the simultaneous presence of MET1-mediated CG methylation and CMT3-mediated CHG methylation at the large intron of the *IBM1* gene is required for proper accumulation of the transcript encoding the functional IBM1 activity. IBM1 targets a large number of CG body-methylated genes for H3K9me2 demethylation, including the gene encoding the ROS1 DNA demethylase as well as ROS1-target genes. Through removal of H3K9me2, the IBM1 activity prevents CHG methylation by protecting genes from CMT3 recruitment. Additionally, ROS1-mediated demethylation contributes to the stabilization of DNA methylation patterns by counteracting small RNA-mediated DNA methylation involving DRM2 at selected loci, including *ROS1*. Loss of CG methylation at the large intron of *IBM1* in *met1* mutants results in lower IBM1 accumulation, and thus to H3K9me2 enrichment at IBM1 target genes. In parallel, *ROS1* transcription is totally turned off in *met1*. These two outcomes of CG methylation erasure are responsible for the appearance of alternative epigenetic patterns at genes: ectopic enrichment in H3K9me2 leads to ectopic CHG body methylation through CMT3 recruitment, while the absence of ROS1 permits DRM2-mediated methylation at CHG and CHH sequence contexts at ROS1-target loci.



heterochromatin? Some proteins require a combinatorial code of epigenetic marks to facilitate their recruitment to chromatin. For instance, CMT3 has been shown to bind *in vitro* to histone H3 peptides that are methylated at both H3K9 and H3K27 positions (Lindroth *et al*, 2004), and in mammals, the Origin recognition complex 2 (Orc2) protein is more efficiently recruited to nucleosomes that exhibit CG methylation in addition to H3K9me3 or H3K27me3 (Bartke *et al*, 2010). Therefore, it is plausible that the efficient recruitment of IBM1 to euchromatin may require DNA methylation and another epigenetic mark. Because we determined that transcriptionally silent genes can potentially be targeted by IBM1 in a *met1* background (Figure 5), this aforementioned additional mark would need to be maintained in euchromatin regions independent of transcription and DNA methylation. However, no epigenetic mark that fulfils these criteria has been described in *Arabidopsis*. Alternatively, IBM1 recruitment may be prevented by the combinatorial presence of DNA methylation and a heterochromatin-specific mark, such that decreasing this mark would result in IBM1 binding and H3K9 demethylation. In support of this scenario, a reduction of the heterochromatin-associated histone modification H4K20me1 in *suvh2* mutants is accompanied by decreased H3K9me2 at heterochromatic chromocentres (Naumann *et al*, 2005). Forward genetics approaches that are designed to identify additional mutants that display ectopic CHG methylation at genes will certainly provide more insights into IBM1 targeting.

The stability of epigenetic patterns along the entire genome is a highly orchestrated process that involves several molecular players that contribute to various epigenetic marks. These factors do not act independently, and in this epigenetic 'symphony', CG methylation may be observed as 'the conductor' because it plays a pivotal role in influencing the homeostasis of many other epigenetic marks. An understanding of the mechanisms that direct genomic CG methylation patterns remains an important challenge for future studies.

## Materials and methods

### Plant material

The *met1-3* (Saze *et al*, 2003), *cmt3-11* (Chan *et al*, 2006), *kyp-7* (Mathieu *et al*, 2005), *nrdp2a-2* (Onodera *et al*, 2005), *drm1-2 drm2-2* (Chan *et al*, 2006), *ddm1-2* (Vongs *et al*, 1993), *vim1 vim2 vim3* (Woo *et al*, 2008), *nrdp1a-4* (SALK\_083051) and *ibm1-4* (Saze *et al*, 2008) mutants were previously described. The *met1-3 cmt3-11* double mutant, and the *kyp-4 suvh5-2 suvh6-1* triple mutant were a kind gift from S Jacobsen and J Bender, respectively. All of the mutants are in the Columbia (Col-0) genetic background. The plants were grown in soil under long-day conditions (16 h light/8 h dark) at 22°C. For the 5-aza-dC treatment, seeds were germinated and grown for 9 days on 1/2 Murashige and Skoog medium that contained 4 µM 5-aza-dC (Sigma) as previously described (Mathieu *et al*, 2007).

### RNA analyses

Total RNA was extracted from immature flower buds (or from leaves for Figure 1C and E) using the TRI Reagent (Sigma). For the RNA gel blot analyses, poly(A) + RNA was purified from 18 µg of total RNA using the Oligotex Kit (Qiagen), and the entirety of the resultant samples was loaded onto the gel. The IBM1 probe was amplified from Col-0 genomic DNA using IBM1-F10 and IBM1-R12 primers (Supplementary Table 1) and was labelled with [ $\alpha$ -<sup>32</sup>P]dCTP using random hexamer priming (Megaprime DNA labeling system, GE Healthcare). The RT-PCR reactions were performed using the OneStep RT-PCR Kit (Qiagen); the amplification of *ACTIN2* or *18S*

*rRNA* was used as an internal control for the RNA amount. A derived-CAPS (dCAPS) assay was used to distinguish allele-specific IBM1-L transcript accumulation in Ler-0  $\times$  *met1-3*, Ler-0  $\times$  *cmt3-11* and Ler-0  $\times$  *kyp-7* F1 hybrids. The IBM1\_polyR1 primer introduces a *Rsa* I restriction site specifically on Ler-0-derived alleles. The primers that were used are listed in Supplementary Table 1.

### Transgene construction and transformation

The IBM1-L transgene—which consisted of 1124 bp of IBM1 promoter sequence fused to the full-length IBM1-L cDNA and 382 bp of sequence downstream of the IBM1-L stop codon—was cloned into the pHm43GW and pKm43GW binary vectors (Karimi *et al*, 2005) using the MultiSite Gateway<sup>®</sup> Three-Fragment Vector Construction technology (Invitrogen). The resultant IBM1-L transgenes (pHm43GW:IBM1-L and pKm43GW:IBM1-L) contain an additional *Bsr*GI restriction site that was used to distinguish them from the endogenous IBM1 gene in the RT-PCR assays. The transgenes were introduced into *Agrobacterium* strain C58C1 and were used to transform plants that were heterozygous for the *met1-3* (pKm43GW:IBM1-L) or *ibm1-4* (pHm43GW:IBM1-L) mutations by floral dipping. The T1 generation transformants were genotyped for the respective mutation, and the homozygous mutants were analysed in comparison with their corresponding homozygous siblings that lacked the IBM1-L transgene.

### DNA methylation analyses

Bisulphite conversion of genomic DNA was performed as previously described (Mirouze *et al*, 2009) using the Epitect Kit (Qiagen). The PCR products were cloned using the pGEM-Teasy vector (Promega). For each sample, 8–12 clones were sequenced. The sequencing data were visualized using the KisMeth software program (Gruntman *et al*, 2008), and the methylation percentage was calculated for each cytosine sequence context (CG, CHG and CHH). The primers that were used are listed in Supplementary Table 1.

For the IBM1 methylation analysis in the F1 hybrids, the intronic IBM1 methylated zone was amplified from bisulphite-treated DNA by nested-PCR with the IBM1-BS-F/IBM1-BS-R and IBM1-BS-Fbis/IBM1-BS-R primer sets (Supplementary Table 1), which facilitated the specific amplification of the Col-0 alleles. The PCR products were digested with *TaqI* (5'-TCGA-3'; 3 sites) and *HpyCH4V* (5'-TGCA-3'; 1 site) to diagnose the CG and CHG methylation, respectively. Methylation at the BNS locus was analysed by the double digestion of 30 ng of genomic DNA with *Bgl*II and *Eco*RI in a 30-µl reaction volume. The control 'undigested' samples were digested with *Eco*RI alone. One microlitre of each sample was amplified by PCR with the BNS-F2 and BNS-R3 primers (Supplementary Table 1) (Saze and Kakutani, 2007).

### Chromatin immunoprecipitation

Leaves of 4-week-old seedlings (0.1 g) were harvested and chromatin was crosslinked and prepared as previously described (Mathieu *et al*, 2005). ChIP was performed with a mouse monoclonal antibody specific for dimethyl H3K9 (Abcam Ab1220) or a negative control IgG from rabbit (Diagenode) using the LowCell# ChIP kit (Diagenode) following the manufacturer's instructions with some modifications. Chromatin samples were pre-cleared with Protein A-coated paramagnetic beads and immunoprecipitated with antibody overnight. Immune complexes were then collected with Protein A-coated paramagnetic beads. ChIP experiments were performed in three independent biological replicates. ChIP samples were amplified in duplicates or triplicates and target amplifications were normalized for amplification of *TUBULIN 8* (*met1* and *met1::IBM1-L* samples) or for amplification of the *Ta3* element (*ibm1* samples), for which H3K9me2 patterns are unaltered in *ibm1* (Saze *et al*, 2008; Inagaki *et al*, 2010) using previously described procedures (Mathieu *et al*, 2005). The primers that were used are listed in Supplementary Table 1.

### Immunocytology

The immunodetection of H3K9me2 was performed with young rosette leaves as previously described (Mathieu *et al*, 2005) using a mouse monoclonal antibody specific for dimethyl H3K9 (Abcam Ab1220, dilution 1:100). The nuclei were visualized with a Zeiss Axio Imager Z.1 microscope that was equipped with the Zeiss AxioCam MRm camera system and the Zeiss Axiovision software program.



# Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

# Acknowledgements

We thank Tetsuji Kakutani, Hidetoshi Saze, Steve Jacobsen, Hervé Vaucheret and Judith Bender for providing Arabidopsis seeds; Ryan Lister for providing the list of methylated cytosines in the WT and *met1-3* mutant; Huan Shu for technical advice; Marie-Noëlle Pouch Pélissier for assistance; and Charles White and Yoko Ikeda for helpful comments. The work in the Mathieu laboratory was supported by the CNRS, the INSERM, the Université

Blaise Pascal, the Université d'Auvergne, and by the European Community's Seventh Framework Programme (FP7 2007/2013) through a Starting Independent Researcher Grant from the European Research Council to OM (I2ST- Grant agreement #260742).

*Author contributions:* MR, TP and OM conceived the study. MR, ZK, TP and OM performed the experiments and analysed the data. OM supervised the work and wrote the paper together with MR.

# Conflict of interest

The authors declare that they have no conflict of interest.

# References

- Bartke T, Vermeulen M, Xhemalce B, Robson SC, Mann M, Kouzarides T (2010) Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**: 470–484
- Baumbusch LO, Thorstensen T, Krauss V, Fischer A, Naumann K, Assalkhou R, Schulz I, Reuter G, Aalen RB (2001) The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Res* **29**: 4319–4333
- Bell AC, Felsenfeld G (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**: 482–485
- Bender J (2004) DNA methylation and epigenetics. *Annu Rev Plant Biol* **55**: 41–68
- Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. *PLoS One* **3**: e3156
- Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* **16**: 6–21
- Bird AP, Wolffe AP (1999) Methylation-induced repression—belts, braces, and chromatin. *Cell* **99**: 451–454
- Cao X, Aufsatz W, Zilberman D, Mette MF, Huang MS, Matzke M, Jacobsen SE (2003) Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr Biol* **13**: 2212–2217
- Cao X, Jacobsen SE (2002) Role of the arabidopsis DRM methyltransferases in *de novo* DNA methylation and gene silencing. *Curr Biol* **12**: 1138–1144
- Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA (2002) Selection for short introns in highly expressed genes. *Nat Genet* **31**: 415–418
- Chan SW, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in Arabidopsis thaliana. *Nat Rev Genet* **6**: 351–360
- Chan SW, Henderson IR, Zhang X, Shah G, Chien JS, Jacobsen SE (2006) RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in Arabidopsis. *PLoS Genet* **2**: e83
- Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, Casero D, Bernal M, Huijser P, Clark AT, Kramer U, Merchant SS, Zhang X, Jacobsen SE, Pellegrini M (2010) Relationship between nucleosome positioning and DNA methylation. *Nature* **466**: 388–392
- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**: 215–219
- Ebbs ML, Bartee L, Bender J (2005) H3 lysine 9 methylation is maintained on a transcribed inverted repeat by combined action of SUVH6 and SUVH4 methyltransferases. *Mol Cell Biol* **25**: 10507–10515
- Ebbs ML, Bender J (2006) Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. *Plant Cell* **18**: 1166–1176
- Feng S, Jacobsen SE (2011) Epigenetic modifications in plants: an evolutionary perspective. *Curr Opin Plant Biol* **14**: 179–186
- Grewal SI, Jia S (2007) Heterochromatin revisited. *Nat Rev Genet* **8**: 35–46
- Gromak N, West S, Proudfoot NJ (2006) Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol Cell Biol* **26**: 3986–3996
- Grunman E, Qi Y, Slotkin RK, Roeder T, Martienssen RA, Sachidanandam R (2008) Kismeth: analyzer of plant methylation states through bisulfite sequencing. *BMC Bioinformatics* **9**: 371
- Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405**: 486–489
- Heger P, Marin B, Schierenberg E (2009) Loss of the insulator protein CTCF during nematode evolution. *BMC Mol Biol* **10**: 84
- Huetzel B, Kanno T, Daxinger L, Aufsatz W, Matzke AJ, Matzke M (2006) Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. *EMBO J* **25**: 2828–2836
- Inagaki S, Miura-Kamio A, Nakamura Y, Lu F, Cui X, Cao X, Kimura H, Saze H, Kakutani T (2010) Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome. *EMBO J* **29**: 3496–3506
- Jackson JP, Johnson L, Jasencakova Z, Zhang X, PerezBurgos L, Singh PB, Cheng X, Schubert I, Jenuwein T, Jacobsen SE (2004) Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. *Chromosoma* **112**: 308–315
- Jackson JP, Lindroth AM, Cao X, Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416**: 556–560
- Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr Biol* **17**: 379–384
- Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, Riddle NC, Verbsky ML, Richards EJ (2003) Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics* **163**: 1109–1122
- Karimi M, De Meyer B, Hilson P (2005) Modular cloning in plant cells. *Trends Plant Sci* **10**: 103–105
- Klose RJ, Kallin EM, Zhang Y (2006) JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet* **7**: 715–727
- Kraft E, Bostick M, Jacobsen SE, Callis J (2008) ORTHO/VIM proteins that regulate DNA methylation are functional ubiquitin E3 ligases. *Plant J* **56**: 704–715
- Lai AY, Fatemi M, Dhasarathy A, Malone C, Sobol SE, Geigerman C, Jaye DL, Mav D, Shah R, Li L, Wade PA (2010) DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas. *J Exp Med* **207**: 1939–1950
- Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11**: 204–220
- Lindroth AM, Shultis D, Jasencakova Z, Fuchs J, Johnson L, Schubert D, Patnaik D, Pradhan S, Goodrich J, Schubert I, Jenuwein T, Khorasanizadeh S, Jacobsen SE (2004) Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. *EMBO J* **23**: 4286–4296
- Lippman Z, May B, Yordan C, Singer T, Martienssen R (2003) Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol* **1**: E67
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133**: 523–536

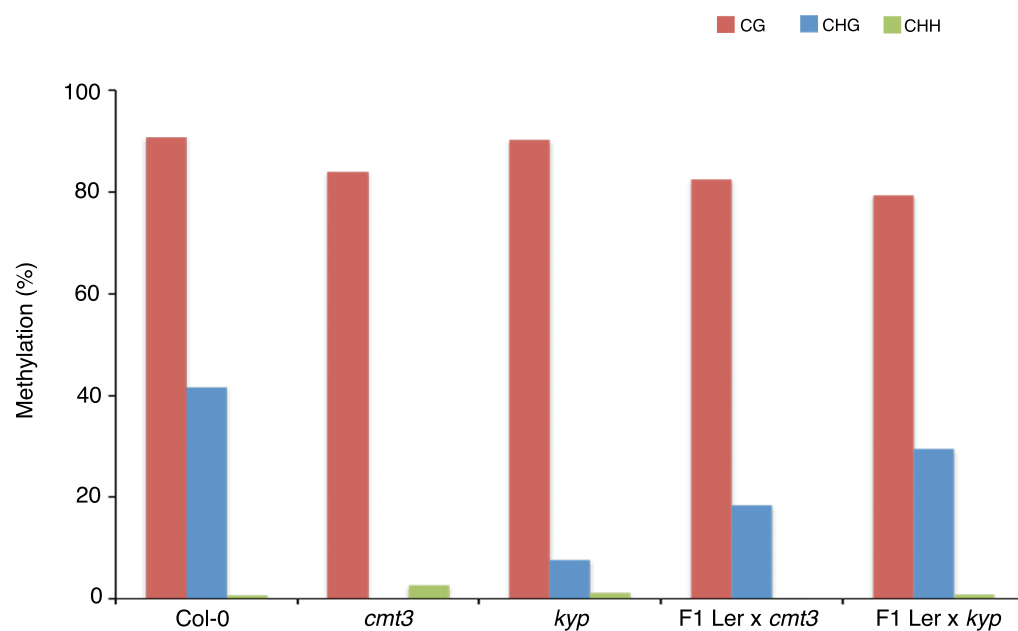


- Lobanenkov VV, Nicolas RH, Adler VV, Paterson H, Klenova EM, Polotskaja AV, Goodwin GH (1990) A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene* **5**: 1743–1753
- Mason PB, Struhl K (2005) Distinction and relationship between elongation rate and processivity of RNA polymerase II *in vivo*. *Mol Cell* **17**: 831–840
- Mathieu O, Probst AV, Paszkowski J (2005) Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in *Arabidopsis*. *EMBO J* **24**: 2783–2791
- Mathieu O, Reinders J, Caikovski M, Smathajitt C, Paszkowski J (2007) Transgenerational stability of the *Arabidopsis* epigenome is coordinated by CG methylation. *Cell* **130**: 851–862
- Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, Cao J, Weigel D, Paszkowski J, Mathieu O (2009) Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature* **461**: 427–430
- Miura A, Nakamura M, Inagaki S, Kobayashi A, Saze H, Kakutani T (2009) An *Arabidopsis* jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. *EMBO J* **28**: 1078–1086
- Naumann K, Fischer A, Hofmann I, Krauss V, Phalke S, Irmeler K, Hause G, Aurich AC, Dorn R, Jenuwein T, Reuter G (2005) Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in *Arabidopsis*. *EMBO J* **24**: 1418–1429
- Onodera Y, Haag JR, Ream T, Costa Nunes P, Pontes O, Pikaard CS (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**: 613–622
- Peng M, Cui Y, Bi YM, Rothstein SJ (2006) AtMBD9: a protein with a methyl-CpG-binding domain regulates flowering time and shoot branching in *Arabidopsis*. *Plant J* **46**: 282–296
- Penterman J, Uzawa R, Fischer RL (2007) Genetic interactions between DNA demethylation and methylation in *Arabidopsis*. *Plant Physiol* **145**: 1549–1557
- Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Theiler G, Chollat D, Descombes P, Paszkowski J (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**: 593–599
- Reinders J, Delucinge Vivier C, Theiler G, Chollat D, Descombes P, Paszkowski J (2008) Genome-wide, high-resolution DNA methylation profiling using bisulfite-mediated cytosine conversion. *Genome Res* **18**: 469–476
- Rodriguez C, Borgel J, Court F, Cathala G, Forne T, Piette J (2010) CTCF is a DNA methylation-sensitive positive regulator of the INK/ARF locus. *Biochem Biophys Res Commun* **392**: 129–134
- Rountree MR, Selker EU (2010) DNA methylation and the formation of heterochromatin in *Neurospora crassa*. *Heredity (Edinb)* **105**: 38–44
- Saze H, Kakutani T (2007) Heritable epigenetic mutation of a transposon-flanked *Arabidopsis* gene due to lack of the chromatin-remodeling factor DDM1. *EMBO J* **26**: 3641–3652
- Saze H, Mittelsten Scheid O, Paszkowski J (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat Genet* **34**: 65–69
- Saze H, Shiraishi A, Miura A, Kakutani T (2008) Control of genic DNA methylation by a jmjC domain-containing protein in *Arabidopsis thaliana*. *Science* **319**: 462–465
- Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**: 74–79
- Soppe WJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, Peeters AJ (2000) The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol Cell* **6**: 791–802
- Soppe WJ, Jasencakova Z, Houben A, Kakutani T, Meister A, Huang MS, Jacobsen SE, Schubert I, Fransz PF (2002) DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in *Arabidopsis*. *EMBO J* **21**: 6549–6559
- Tariq M, Saze H, Probst AV, Lichota J, Habu Y, Paszkowski J (2003) Erasure of CpG methylation in *Arabidopsis* alters patterns of histone H3 methylation in heterochromatin. *Proc Natl Acad Sci USA* **100**: 8823–8827
- The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815
- Tran RK, Henikoff JG, Zilberman D, Ditt RF, Jacobsen SE, Henikoff S (2005) DNA methylation profiling identifies CG methylation clusters in *Arabidopsis* genes. *Curr Biol* **15**: 154–159
- Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P, Zhang Y (2006) Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**: 811–816
- Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, Miura A, Kakutani T (2009) Bursts of retrotransposition reproduced in *Arabidopsis*. *Nature* **461**: 423–426
- Unoki M, Nakamura Y (2003) Methylation at CpG islands in intron 1 of *EGR2* confers enhancer-like activity. *FEBS Lett* **554**: 67–72
- Uptain SM, Kane CM, Chamberlin MJ (1997) Basic mechanisms of transcript elongation and its regulation. *Annu Rev Biochem* **66**: 117–172
- Vongs A, Kakutani T, Martienssen RA, Richards EJ (1993) *Arabidopsis thaliana* DNA methylation mutants. *Science* **260**: 1926–1928
- Wang BB, Brendel V (2006) Genomewide comparative analysis of alternative splicing in plants. *Proc Natl Acad Sci USA* **103**: 7175–7180
- Woo HR, Dittmer TA, Richards EJ (2008) Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in *Arabidopsis*. *PLoS Genet* **4**: e1000156
- Woo HR, Pontes O, Pikaard CS, Richards EJ (2007) VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. *Genes Dev* **21**: 267–277
- Xiao W, Gehring M, Choi Y, Margossian L, Pu H, Harada JJ, Goldberg RB, Pennell RI, Fischer RL (2003) Imprinting of the MEA Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Dev Cell* **5**: 891–901
- Yaish MW, Peng M, Rothstein SJ (2009) AtMBD9 modulates *Arabidopsis* development through the dual epigenetic pathways of DNA methylation and histone acetylation. *Plant J* **59**: 123–135
- Yamane K, Toumazou C, Tsukada Y, Erdjument-Bromage H, Tempst P, Wong J, Zhang Y (2006) JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* **125**: 483–495
- Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE (2009) Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol* **10**: R62
- Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker JR (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126**: 1189–1201
- Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S (2008) Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456**: 125–129
- Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* **39**: 61–69



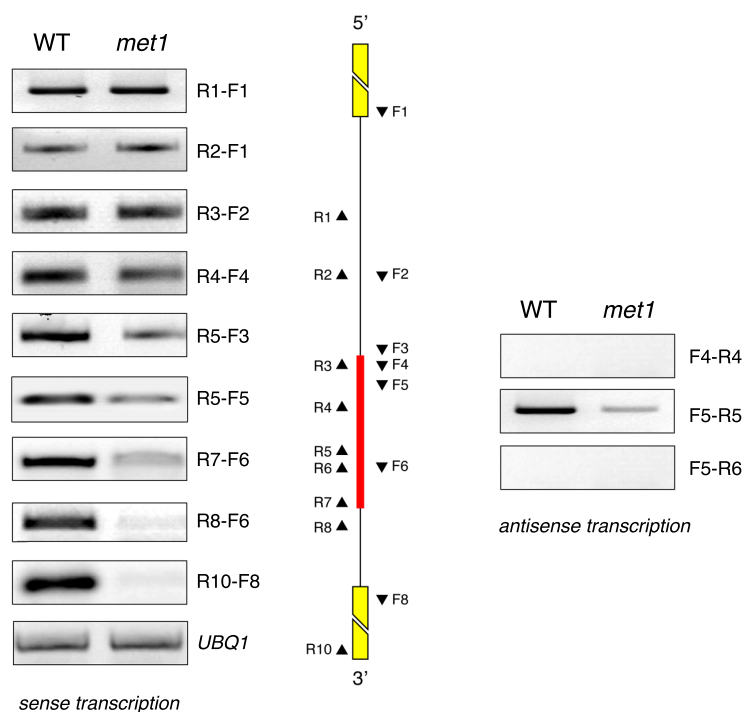


# Supplemental Data



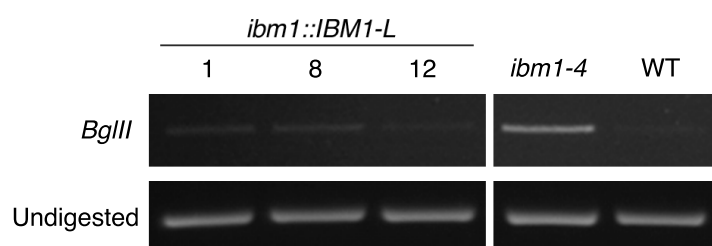
**Supplementary Figure S1.** Bisulfite sequencing analysis of cytosine methylation at the methylated zone of the large *IBM1* intron in wild-type, *cmt3* and *kyp* plants, and of the *cmt3*-derived and the *kyp*-derived alleles in Ler x *cmt3* and Ler x *kyp* F1 hybrids, respectively. Proportions of methylated cytosines at CG (red), CHG (blue) and CHH (green) sites are indicated as percentages.





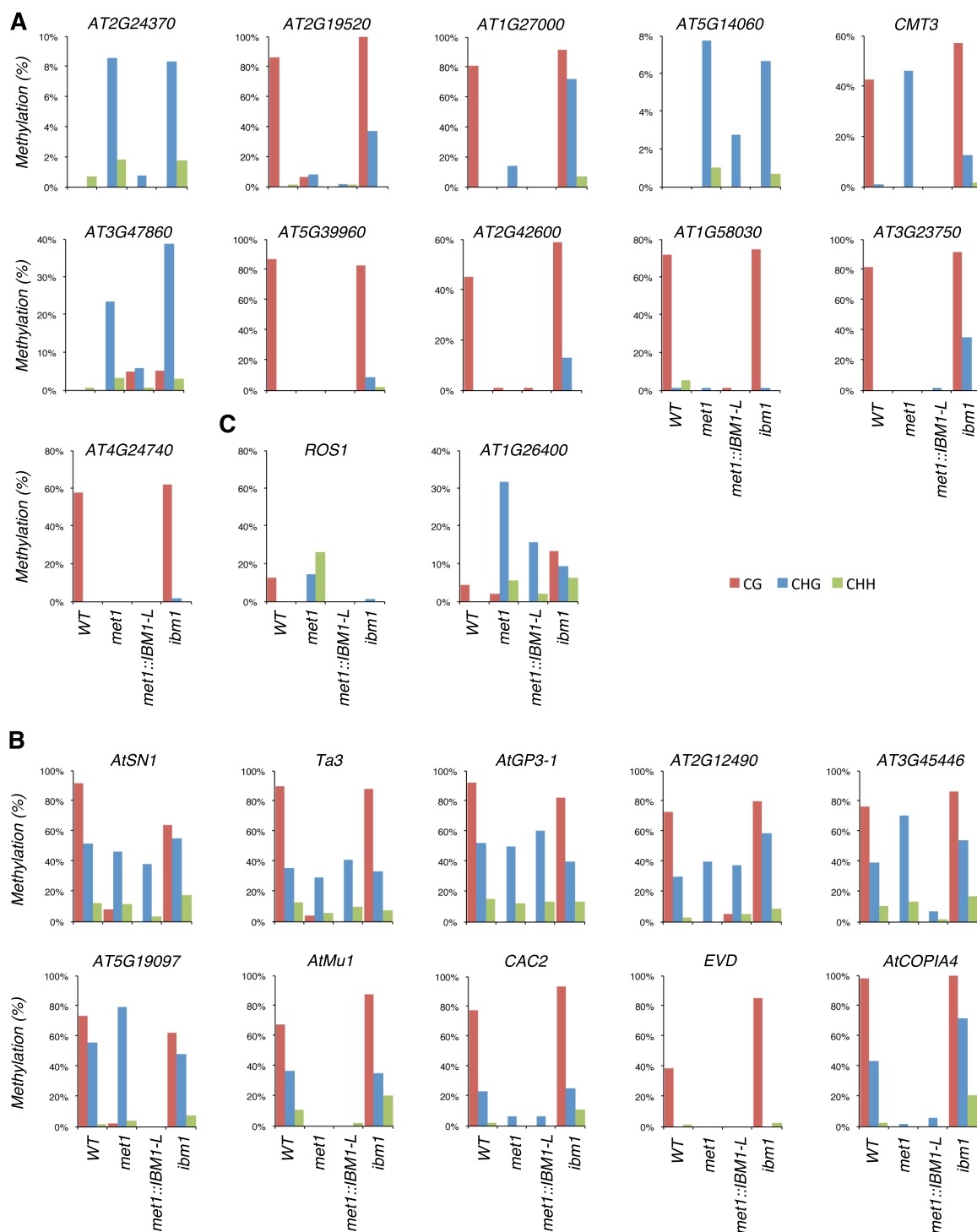
**Supplementary Figure S2.** RT-PCR analysis of sense (left panel) and antisense (right) *IBM1-L* transcription in wild-type and *met1-3* plants, using primer sets designed along the length of the large intron of the *IBM1* gene (middle). Primer positions, intron, exons and the methylated intronic region are indicated by black triangles, a black line, yellow boxes and a red bar, respectively. The amplification of *UBQ1* was used to normalize the RNA template levels.





**Supplementary Figure S3.** CHG methylation at the *BNS* locus in 3 independent *ibm1::IBM1-L* lines and in *ibm1* and wild-type plants. DNA methylation was analyzed by restriction digestion with the methylation-sensitive enzyme *BglII* followed by PCR with BNS-F2/R3 primers (Supplementary Table 1).

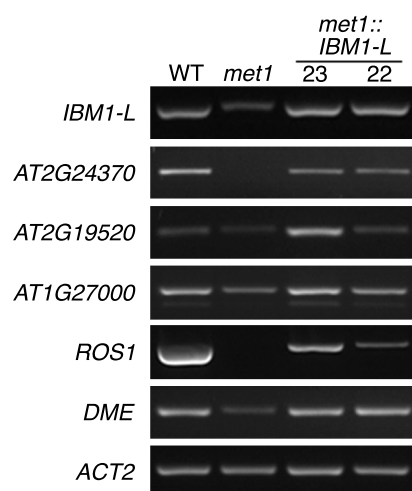




**Supplementary Figure S4.** Bisulfite sequencing analysis of the DNA methylation patterns at eleven genes (A), ten transposable elements (B) and at the *ROS1* gene and the *AT1G26400* *ROS1* target locus (C), in wild-type, *met1*, *met1::IBM1-L* and *ibm1* plants. Proportions of methylated cytosines at CG (red), CHG (blue) and CHH (green) sites are indicated as percentages.







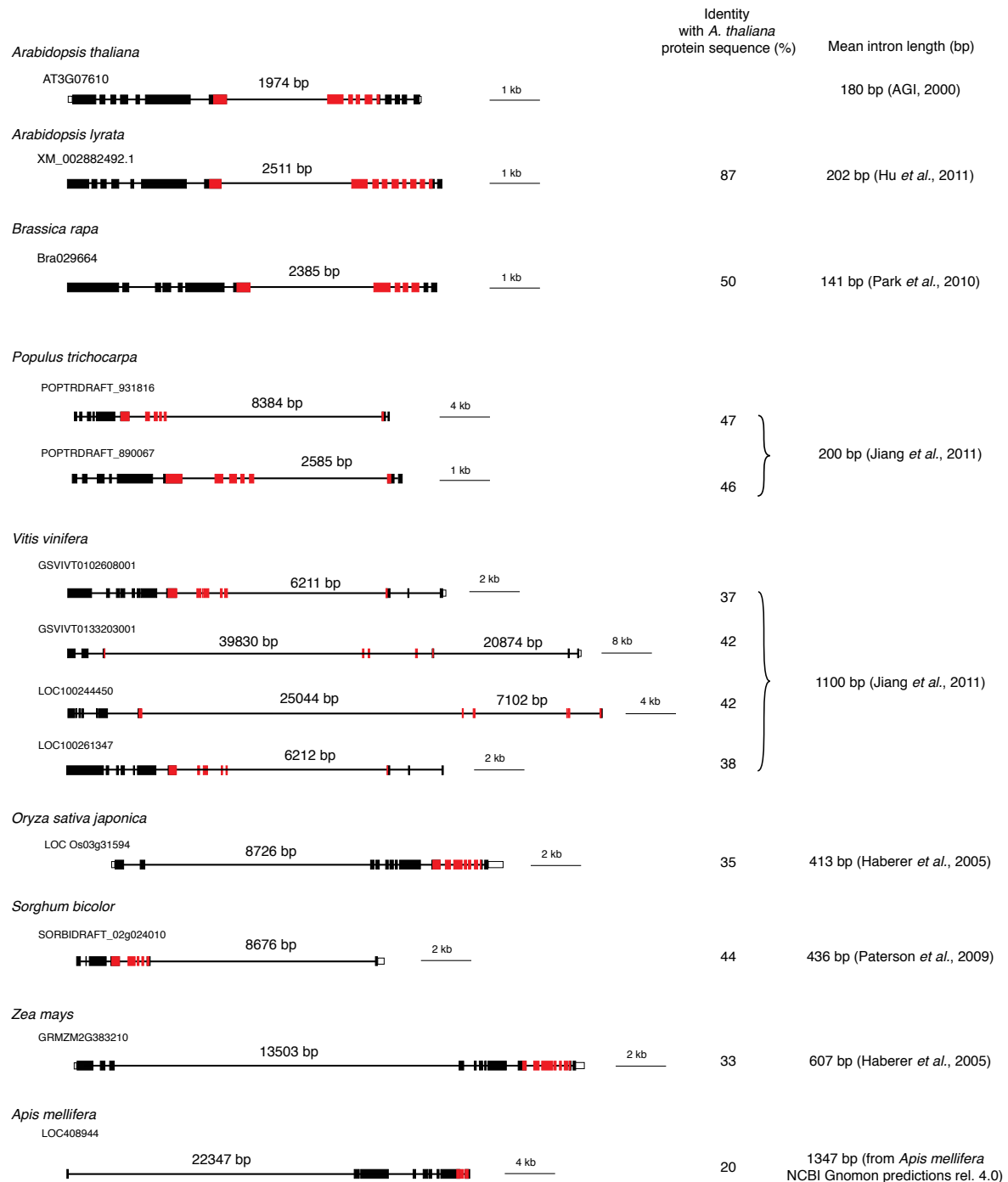
**Supplementary Figure S5.** Transcription analysis in 2 independent *met1::IBM1-L* lines. Representative images of RT-PCR analyses at *IBM1-L*, *AT2G24370*, *AT2G19520* and *AT1G27000*, *ROS1* and *DME* genes.





**Supplementary Figure S6.** Abnormal flower phenotypes observed in some *met1-3* lines. Images of a whole inflorescence shoot (A) and magnification of two abnormal siliques (B & C).





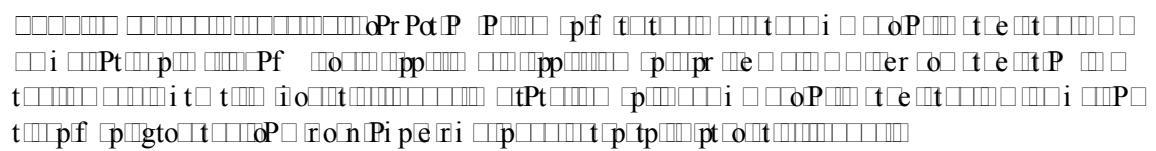
**Supplementary Figure S7.** Orthologs of the *A. thaliana* IBM1 gene containing a large intron.

Boxes represent exons, with coding regions in black and untranslated regions in white. The region encoding the jmjC domain is shown in red (Saze *et al.*, 2008). Length of the large introns and scale bars are indicated for each gene. The mean length of introns in each organism and the percentage of identity of each orthologous protein with Arabidopsis IBM1 protein are indicated.

Orthologous genes were identified by blastp of the Arabidopsis IBM1 protein sequence against non-redundant protein sequences of the NCBI database (<http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>). Additional genes annotated as orthologs of the IBM1 gene in the Gramene database were selected (<http://www.gramene.org/>).

The percentage of identity of each orthologous protein with the Arabidopsis IBM1 protein sequence were determined by alignment of orthologous protein sequences using the ClustalW2 program (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>).









Locus	Primer	Sequence (5'-3')	Methods
<b>18S rRNA</b>	18S-F2	TCCTATTGTGTGGCTTCGG	RT-PCR
	18S-R	CACGTAGCTAGTTAGCAGGCTG	
<b>ACT2</b>	ACT2-F <sup>1</sup>	CTAAGCTCTCAAGATCAAAGGC	RT-PCR
<b>(AT3G18780)</b>	ACT2-R <sup>1</sup>	AACATTGCAAAGAGTTTCAAGG	
<b>AT1G26400</b>	At1g26400-RT-F1	GATCTTAGCTTCATCCACGTTCAA	RT-PCR, ChIP
	At1g26400-RT-R1	CGTCGACTCCAGGAGAATT	
	At1g26400-BS-F	CTTAAACTACRCTTTTAATACTTAACTAAT	Bisulfite sequencing
	At1g26400-BS-R	GGGGAATTTAAAGAGATTGATGGATGTTAAAG	
<b>AT1G27000</b>	At1g27000-F	TGTGAATTCTGCTGTTTGTG	RT-PCR
	At1g27000-R	ACAACCACATCACGTTCCAA	
	AT1G27000-BS-F	CAAAATTATRCATRAATTCCAAATCCATCTCAAA	Bisulfite sequencing, ChIP (with AT1G27000-BS-Fbis/R)
	AT1G27000-BS-Fbis	CATAACCACAACCACATCACRTTCCAAAAAC	
	AT1G27000-BS-R	TTGAAAAGAGATTTTGTGTTGTAAGGGAAT	
<b>AT1G58030</b>	AT1G58030-F	ATTTACAAAGGCCGTGGACA	RT-PCR, ChIP
	AT1G58030-R	TGTATCGGTGTGGCTGCTTA	
	AT1G58030-BS-F	CCCRTRTRTCCCCTCCTACCTATCATCT	Bisulfite sequencing
	AT1G58030-BS-R	GATTTGTGTTAYATATATAAYAAAAGAGATG	
	AT1G58030-BS-Fbis	CCTACCTATCATCTRCATTCTCATCAACAT	
<b>AT2G12490</b>	copia66_F	GTGCCCCGTGACCCTGGAT	RT-PCR
	copia66_R	CTGATTGACTAGAACCGAAGAA	
	AT2G12490-BS-F	CAACACAACRATCAARCTTCAATTC	Bisulfite sequencing
	AT2G12490-BS-Fbis	ATCRATCAARAACATAATCAACAAAAA	
	AT2G12490-BS-R	TGAAGGTGATYAAAYGGAAGTAAGT	
<b>AT2G19520</b>	At2g19520-F	AAACAAGGGAACCTAGAAGCAA	RT-PCR
	At2g19520-R	TCTCTTAACCCACTGACACTATCTTG	
	AT2G19520-BS-F	TATRAARTATATCTCTTAACCCACT	Bisulfite sequencing
	AT2G19520-BS-Fbis	CTCTTAACCCACTRACACTATCTTA	
	AT2G19520-BS-R	GGGAAYTTAGAAGYAAGTAAATGAGAAAGT	
<b>AT2G24370</b>	AT2G24370_F1	TGCGGATGTTTGTCTACGAGTTC	RT-PCR, ChIP
	AT2G24370_R1	CATTGGTGGTTTTGCTGTTATCA	
	AT2G24370-F	ATTTTACAAATTTTAACTTTCTCCACAA	Bisulfite sequencing
	AT2G24370-Fbis	ACACTCTATTRCATCTCCTTATACTAC	
	AT2G24370-R	ACTAGAGAAAGGAACCTTGGTTGATTTA	
<b>AT2G42600</b>	AT2G42600-F	AAAACCGAAGAGGTTGCAGA	RT-PCR, ChIP
	AT2G42600-R	ACCCGATCATGACTTCTTGC	
	AT2G42600-BS-F	CCTCRTCCAACRRTCCACCTCTTCC	Bisulfite sequencing
	AT2G42600-BS-R	GAATGGTYGGAGGATAAAAGAYAGGAATGG	
	AT2G42600-BS-Fbis	CTCCRTATTCTTTTCCACCTTCAC	
<b>AT3G23750</b>	AT3G23750-F	TGGATGTGTATGCATTGTTG	RT-PCR, ChIP
	AT3G23750-R	ATGTCGGGTCTTTGTTGAGG	
	AT3G23750-BS-F	TTACATCRATTCRAAACTCTCTTC	Bisulfite sequencing
	AT3G23750-BS-R	ATTGGTTGYAYTGGATATTGATGTTATA	
	AT3G23750-BS-Rbis	GGAAGAGTAAYGAYGAAAGTGGATGTGTATG	
<b>AT3G45446</b>	AT3G45446-F	CCATTGTACTCCCAACACCA	RT-PCR, ChIP
	AT3G45446-R	TGAGAATTATGGCAACAAGGTG	
	AT3G45446-BS-F	TACTCCCAACACCATATCAATACCTT	Bisulfite sequencing
	AT3G45446-BS-R	GTGGATYAGAAAGGTGAGAATTATGG	
	AT3G45446-BS-Fbis	CCATCCACATTAATTTTCTCCCATC	
<b>AT3G47860</b>	AT3G47860-F	TCAGAAGAGTGTGTTGATTCTGGA	RT-PCR, ChIP
	AT3G47860-R	TTGGCACAATTTGGCTATGA	
	AT3G47860-BS-F	CTCTTRCACTRATCTTATATAAACCTAAT	Bisulfite sequencing
	AT3G47860-BS-R	AYTIAAGATYYGATGTGGGATAAAAT	
	AT3G47860-BS-Rbis	GTAYTTTGGTTTYTYTGTGGAGTGTATG	
<b>AT4G24740</b>	AT4G24740-F	TGGTCAAGTGCTAGAATGTTGG	RT-PCR, ChIP
	AT4G24740-R	ATGGTTACGATAGTCAAACCAGTTC	
	AT4G24740-BS-F	ACTACACACATRCATACATTCCAA	Bisulfite sequencing
	AT4G24740-BS-R	GTTGGGATAGGGAGAGGAAGGAAATGGTGG	
	AT4G24740-BS-Rbis	GTGAGGYTGYTATGATTGAAATTGAAATG	



<b>AT5G14060</b>	AT5G14060-F	GCAATGGAGTGAATGTGCAG	RT-PCR
	AT5G14060-R	CAACAACAACAACAATGGA	
	AT5G14060-BS-F	TCTTRACACATTCTTARCACTAACTAC	Bisulfite sequencing, ChIP (with AT5G14060-BS-F/Rbis)
	AT5G14060-BS-R	TTTGATAGATTGTTTTTTGTGAGGGT	
	AT5G14060-BS-Rbis	GGGTYATATGTGGTATTGGTTAAAGGG	
<b>AT5G19097</b>	AT5G19097-F	CGAAGCCACAACATTTCCTT	RT-PCR, ChIP
	AT5G19097-R	CAGCCACTGCGTCTTTTACA	
	AT5G19097-BS-F	CTRAARCATCAATCCACAATCATTCA	Bisulfite sequencing
	AT5G19097-BS-R	GAAAGAATYTTTTAAAYTGAGAGATTAGG	
	AT5G19097-BS-Fbis	CTTTCRAARCCACAACATTTCCTT	
<b>AT5G39960</b>	AT5G39960-F	GAGGGAAGAGGACGTATGGA	RT-PCR, ChIP
	AT5G39960-R	TAGGAGTCCCGCCTAAATCA	
	AT5G39960-BS-F	CTRTTTCRCCTTCCACCACCCT	Bisulfite sequencing
	AT5G39960-BS-R	GGAAGAGGAYGTATGGAAGTAATGAAAGAAG	
	AT5G39960-BS-Fbis	TCCRCCTAAATCAAAATCTTCCTCAAT	
<b>AtCOPIA4</b> <b>(AT4G16870)</b>	AtCOPIA4-F <sup>2</sup> (JP3067)	CTTGTTGTCTTCCCGTGT	RT-PCR, ChIP
	AtCOPIA4-R <sup>2</sup> (JP3087)	TGACGAAGAGCGTACCTGTG	
	AtCOPIA4-BS-R <sup>3</sup> (JP3100)	GGTTGTYTGTGTTTTTATGGTTYAGATTTATA	Bisulfite sequencing
	AtCOPIA4-BS-F <sup>3</sup> (JP3101)	ATAACTRAACCACARATTCARACCCATTTTCATT	
<b>AtGP3-1</b> <b>(AT3G11970)</b>	AtGP3-1-F	AAGCATTAATAGTACAAAGTTCCCC	RT-PCR
	AtGP3-1-R	TTCTCTTGGAAAGTTGCAGAC	
	AtGP3-1_BS-F <sup>4</sup>	ATGYTAGTGTATYTTAAATAGTGATYTTG	Bisulfite sequencing, ChIP (with AtGP3-1_BS-Fbis/R)
	AtGP3-1_BS-R <sup>4</sup>	AACRTTTACTTTCTTTACCTTTTATATAA	
	AtGP3-1_BS-Fbis	GGTTYAGGATAGATTATTAAYGGGGGAA	
<b>AtMu1</b> <b>(AT4G08680)</b>	AtMu1-F <sup>5</sup>	CCGAGAACTGGTTGTGGTTT	RT-PCR
	AtMu1-R <sup>5</sup>	GCTCTTGCTTTGGTGATGGT	
	AtMu1-BS-F1 <sup>6</sup>	AATTTTATGGAATGAAGTTATATG	Bisulfite sequencing, ChIP (with AtMu1-BS-F1/Rbis)
	AtMu1-BS-R	AAATTTAARCCCTCTTCTCATACA	
	AtMu1-BS-Rbis <sup>6</sup>	TTCTCATACARTRCTTCAATTT	
<b>AtSN1</b> <b>(AT3G44000-AT3G44005)</b>	ATS15 <sup>7</sup>	ACCAACGTGTTGTTGGCCAGTGGTAAATC	RT-PCR, ChIP
	ATS1N1-F4 <sup>7</sup>	AAAATAAGTGGTGGTTGTACAAGC	
	AtSN1-BS-F	CTCTCAAACAAACCCCTTAAAC	Bisulfite sequencing
	AtSN1-BS-R	GTTGAGATGGTGRTRTTGGAAAGG	
	AtSN1-BS-R1	GGAAGTTTGTATAGTTTAAARATAAG	
<b>BNS</b> <b>(AT1G73177)</b>	BNS-F2 <sup>8</sup>	GCTAGAGGTTTTAGTTCTCTG	BNS methylation analysis
	BNS-R3 <sup>8</sup>	TTCCCTATGACATTCAAGGTC	
<b>CAC2</b> <b>(AT1G35270-AT1G35280)</b>	CAC2A2-RT-F	TTGGTCCGATAAGAGTTTCGT	RT-PCR
	CAC2A2-RT-R	CGCAGGTTTCTTTCTTCAGC	
	CAC2-F	GCCTAAGTGTGGCGGACTA	ChIP (with CAC2-bsR2)
	CAC2-bsF2 <sup>9</sup>	CATATAAACCCCAAAATCAAATC	
	CAC2-bsR2 <sup>9</sup>	ATGGAAAAGGAGAAGGAGGTAT	Bisulfite sequencing
	CAC2-bsR2bis	TYATTGATAGYAGTAGAGGATAG	
<b>CMT3</b> <b>(AT1G69770)</b>	CMT3-MseI-F	ATGCCAAGGAATCAGTGGTC	RT-PCR
	CMT3-MseI-R	AACTTTTGAAATGGCGTTGTG	
	CMT3-R2	TAATTCCGCAGTCCCACTT	ChIP (with CMT3-BS-Fbis)
	CMT3-BS-F	CCTAATTTTARTCAAAATTRCTAAAAACATC	
	CMT3-BS-Fbis	TCAAAATTTTRCTAAAAACATCAATACTTA	Bisulfite sequencing
	CMT3-BS-R	GGTGTATATGAAYATTGTAGAATATTTGAAG	
<b>DME</b> <b>(AT5G04560)</b>	DME-F	TGAGAGGAAGTTTTCCGCTC	RT-PCR
	DME-R	GTGTCGCTTCTGCTCTATGA	
<b>EVD</b> <b>(AT5G17125)</b>	EVD_RT_F <sup>10</sup>	GACCTGACATTTGTTTTGCG	RT-PCR
	EVD_RT_R <sup>10</sup>	GGTGAAGATATCCGCCAAC	
	EVD_TD	ACTCGTGAGTCCCTCTTCAACGG	ChIP (with EVD-BS-R)
	EVD-BS-F <sup>10</sup>	ACATRATCTTATRCTTRATACCAT	
	EVD-BS-R1 <sup>10</sup>	GTAGAGGAYAAATGTAAATYGTGTGG	Bisulfite sequencing
	EVD-BS-R2 <sup>10</sup>	GTGYGAYAAGATYGTAGGTTATTG	



<b>IBM1</b> (AT3G07610)	IBM1-F8	ACAACAAGTCCAAAATGTTG	RT-PCR (IBM1-L)
	IBM1-R10	TAAACACTCGCTGACATTTTCAGG	
	IBM1-F9	TATGTAGTTTGCTTTGAAG	RT-PCR (IBM1-S)
	IBM1-R11	GTTCTGAAATCATTATAGATGTGC	
	IBM1-F10	TGAGATCCGCAATGGAAAG	probe for Northern blot analysis
	IBM1-R12	CCAAACAATCAACAGCAAC	
	IBM1-F1	TCTGATGCGGTTAGTACCC	
	IBM1-R1	CCAAAAACCGAATGCCAGGCC	
	IBM1-R2	TAAGCAGAATGGATGAAGGC	
	IBM1-F2	AGCCTTCATCCATTCTGCTT	
	IBM1-R3	GCGTAATCTTGCATTCAATGTG	
	IBM1-F3	CAGCATTTTATAAATGTTTGATTG	
	IBM1-F4	ACATTGAATGCAAAGATTACGC	RT-PCR along the length of IBM1 gene, ChIP (with IBM1- F5/R6)
	IBM1-F5	TAAAGTGTTTGTGATTGTTTGA	
	IBM1-F6	ACTTCTTGATATTGTTGGAATG	
	IBM1-R4	CATTATTGATACGACAGACG	
	IBM1-R5	GCGAAATTGATGTTGGTCATT	
	IBM1-R6	ATCATTCCGAACAATATCAAGA	
	IBM1-R7	TGGAAAGAGTCATTGAGGTAG	
	IBM1-R8	CACAAAACCAAAATCTACTAAATTAATATT	
	IBM1_polyR1	TTAGTACTACTCTGTTAACTTAGAGGTA	dCAPS analysis
	IBM1-BS-F	GTTAATYGTYGATATAAYTGGTGTAGTT	Bisulfite sequencing, methylation analysis by restriction digestion
	IBM1-BS-R	CACAAAACCAAAATCTACTAAATTAATATT	
	IBM1-BS-Fbis	TAGATTTTATATGTTGYGTGTA	
<b>ROS1</b> (AT2G36490)	ROS1-F	AGAAGAAATTCCTACCATCA	RT-PCR
	ROS1-R	ACCGTTCTTCGAGGTAATTC	
	ROS1-R2	AGAGGATGCTCGTCTGGAAG	ChIP (with ROS1-F)
	ROS1_BS_F <sup>11</sup>	TAGAGGAGGYGTTTTTTG	Bisulfite sequencing
	ROS1_BS_Fbis <sup>11</sup>	GAGGYGTTTTTTGAGGAT	
	ROS1_BS_R <sup>11</sup>	CATTTTAACTAAAAACRAAAAAA	
<b>Ta3</b> (AT1G37110)	Ta3-middle-F <sup>2</sup> (JP1565)	GATTCTTACTGTAAAGACATGGCATTGAGAGA	RT-PCR
	Ta3-middle-R <sup>2</sup> (JP1566)	TCCAAATTTCTGAGGTGCTTGTAACC	
	Ta3-LTR-F <sup>2</sup> (JP1617)	TAGGGTTCTTAGTTGATCTGTATTGAGCTC	ChIP
	Ta3-LTR-R <sup>2</sup> (JP1618)	TTTGCTCTCAAACTCTCAATTGAAGTTT	
	Ta3-BS-F (JP1615) <sup>3</sup>	GGTTTAAITGTGGTTTAGTGATATTYGGTTTAGT	Bisulfite sequencing
	Ta3-BS-Rbis (JP1616) <sup>3</sup>	AATCAAAAAACRAATAAACCTCRCTCTRATACCACTTATT	
	Ta3-BS-R	CCATTAAACACCTTTATCTCAACTCTA	
<b>TUB8</b> (AT5G23860)	TUB8-F <sup>12</sup>	ATAACCGTTTCAAATCTCTCTCTC	ChIP
	TUB8-R <sup>12</sup>	TGCAAATCGTTCTCTCCTTG	
<b>UBQ1</b> (AT3G52590)	UBQ-FWD (U1) <sup>13</sup>	ACCGGCAAGACCATCACTCT	RT-PCR
	UBQ-Rev (U2) <sup>13</sup>	AGGCCTCAACTGGTTGCTGT	

**Supplementary Table 1.** List of primers.

<sup>1</sup> Vaillant I, Schubert I, Tourmente S, Mathieu O (2006) MOM1 mediates DNA-methylation-independent silencing of repetitive sequences in Arabidopsis. *EMBO J* 7: 1273-1278

<sup>2</sup> Johnson L, Cao X and Jacobsen S (2002) Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr Biol* 12: 1360-1367.

<sup>3</sup> Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, Jacobsen SE (2006) Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet* 38: 721-725

<sup>4</sup> Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, Miura A, Kakutani T (2009) Bursts of retrotransposition reproduced in Arabidopsis. *Nature* 461: 423-426

<sup>5</sup> Singer T, Yordan C, Martienssen RA (2001) Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA methylation (DDM1). *Genes Dev* 15: 591-602



- <sup>6</sup> Gruntman E, Qi Y, Slotkin RK, Roeder T, Martienssen RA, Sachidanandam R (2008) Kismeth: analyzer of plant methylation states through bisulfite sequencing. *BMC Bioinformatics* **9**: 371
- <sup>7</sup> Herr A J, Jensen MB, Dalmay T, Baulcombe DC (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**: 118-120.
- <sup>8</sup> Saze H, Kakutani T (2007) Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1. *EMBO J* **26**: 3641-3652
- <sup>9</sup> Miura A, Nakamura M, Inagaki S, Kobayashi A, Saze H, Kakutani T (2009) An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. *EMBO J* **28**: 1078-1086
- <sup>10</sup> Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, Cao J, Weigel D, Paszkowski J, Mathieu O (2009) Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* **461**: 427-430
- <sup>11</sup> Reinders J, Delucinge Vivier C, Theiler G, Chollet D, Descombes P, Paszkowski J (2008) Genome-wide, high-resolution DNA methylation profiling using bisulfite-mediated cytosine conversion. *Genome Res* **18**: 469-476
- <sup>12</sup> Mathieu O, Probst A, Paszkowski J (2005) Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. *EMBO J* **24**: 2783-2791
- <sup>13</sup> Zhang W, Sun Y, Timofejeva L, Chen C, Grossniklaus U, Ma H (2006) Regulation of Arabidopsis tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. *Development* **133**: 3085-3095





**III. Chapitre 2 : Conséquences immédiates de la rencontre de plantes aux génomes identiques mais aux épigénomes distincts – Dynamique de la méthylation ADN et de la transcription dans les épihybrides F1.**

**ARTICLE**

**“Epigenome confrontation triggers immediate reprogramming of DNA methylation and transposon silencing in Arabidopsis epihybrids”**

Rigal M\*, Becker C, Pélissier T, Pogorelcnik R, Devos J, Ikeda Y, Weigel D, Mathieu O

Article soumis.



### III.A. Introduction

Lors de l'analyse de la méthylation de l'intron du gène *IBM1* dans les plantes F1 *met1* x WT rapportée dans le premier article (Rigal et al., 2012), nous avons observé, de façon très surprenante, une perte de la méthylation CHG sur l'allèle venant du parent *met1*. Nous avons souhaité comprendre la cause de cette déméthylation mais également le comportement global des profils de méthylation ADN lors de la rencontre entre des plantes hypométhylées *met1* et des plantes sauvages. Dans ce but, nous nous sommes tout d'abord tournés vers une technique déjà utilisée chez *A. thaliana*, mais aussi en cellules humaines, la MeDIP, immunoprécipitation de l'ADN méthylé grâce à un anticorps anti-5-meC. Ceci s'est suivi d'un séquençage haut-débit de la fraction non immunoprécipitée (non méthylée), en partant de l'hypothèse qu'une déméthylation telle qu'observée sur *IBM1* en F1 se traduirait par une hausse claire du nombre de reads par rapport aux parents. Cependant l'analyse s'est révélée ardue, en partie en raison de la présence d'un bruit de fond important, variable selon les génotypes des échantillons, et difficilement quantifiable.

Afin de contourner cette difficulté, nous avons séquençé la fraction immunoprécipitée (méthylée) de la MeDIP, comme cela avait déjà été fait, notamment par Zhang *et al.*, en 2006 (Zhang et al. 2006). Toutefois, l'étude des données en résultant a également été très complexe. En effet, la présence de méthylation dans les trois contextes compliquait l'analyse puisque, comme cela a été précédemment démontré, l'anticorps semble avoir un taux d'immunoprécipitation variable selon le contexte dans lequel se trouve la cytosine méthylée et même selon la densité locale en cytosines (Irizarry et al. 2008, Lister et al. 2008, Tomazou et al. 2008). De plus, dans les plantes F1, la présence à la fois d'allèles méthylés sauvages ainsi que d'allèles hypométhylés venant du parent *met1* rend l'analyse des données de séquençage encore plus ardue, puisque nous ne pouvons connaître la contribution de chaque allèle dans le taux global de méthylation observée dans les hybrides F1.

Afin de venir à bout de ces biais, nous avons testé différentes méthodes de calcul tentant de mesurer le biais dans chaque échantillon, et de définir des facteurs de correction appropriés en prenant comme références des loci dont le taux de méthylation dans les parents et/ou en F1 avait été publié et/ou que j'avais analysé moi-même par séquençage après traitement de l'ADN au bisulfite de sodium.



Les études publiées auparavant ayant utilisé cette méthode de MeDIP suivie d'un séquençage génome entier (MeDIP-seq) ou d'une hybridation sur puces de l'ADN immunoprécipité (MeDIP-ChIP) portaient, soit sur des échantillons issus de cellules de Mammifères, ne comportant que, ou en grande majorité, de la méthylation CG (Weber et al. 2005, Keshet et al. 2006, Weber et al. 2007, Lisanti et al. 2012), soit dans des plantes homozygotes - sauvages ou *met1* (Zhang et al. 2006, Zilberman et al. 2007, Lister et al. 2008, Vining et al. 2012).

J'ai ainsi consacré un temps très important à l'analyse de ces données, et même si certains des changements particulièrement conséquents dans les taux de méthylation ADN en F1 ont été confirmés par McrBC, toutes les méthodes de calcul testées offraient une proportion importante de faux-positifs parmi les DMRs mises en évidence.

Ainsi, pour pouvoir obtenir une information complète et fiable, nous avons réalisé durant les derniers mois de ma thèse un séquençage bisulfite du génome entier sur l'ADN des plantes Col-0, *met1* et des plantes F1 *met1* x WT.

L'analyse des données de séquençage a débuté lors de la période de rédaction de mon manuscrit de thèse, par O. Mathieu, R. Pogorelnik, bioinformaticien intégré à l'équipe, et moi-même.

Elle a mené à la rédaction d'une première version préliminaire d'un article intégré au manuscrit de thèse, présenté lors de la soutenance. Les réflexions du jury de thèse lors de la soutenance sur ces données préliminaires nous ont aidé à poursuivre l'analyse et à revoir la structuration du manuscrit de l'article. Les suggestions du jury nous ont mené à réaliser un séquençage des petits ARN (sRNA-seq) dans les mêmes échantillons, pour pouvoir corrélérer taux de méthylation ADN, transcription et taux d'accumulation des siRNA.

L'analyse des données du sRNA-seq a été réalisée par O. Mathieu et R. Pogorelnik.



### III.B. Article « Rigal et al., 2015 »

#### **Epigenome confrontation triggers immediate reprogramming of DNA methylation and transposon silencing in *Arabidopsis* epihybrids**

*Running title:* **Epigenomic shock in *Arabidopsis* F1 epihybrids**

Mélanie Rigal<sup>1,2</sup>, Claude Becker<sup>3</sup>, Thierry Péliissier<sup>1,2</sup>, Romain Pogorelcnik<sup>1,2</sup>, Jane Devos<sup>3</sup>, Yoko Ikeda<sup>4</sup>, Detlef Weigel<sup>3</sup>, Olivier Mathieu<sup>1,2</sup>

<sup>1</sup> Laboratoire Génétique Reproduction et Développement (GReD), CNRS, UMR 6293, INSERM, UMR 1103, F-63001 Clermont-Ferrand, France

<sup>2</sup> Université Clermont Auvergne, Université Blaise Pascal, GReD, BP 10448, F-63000 Clermont-Ferrand, France

<sup>3</sup> Department of Molecular Biology, Max Planck Institute for Developmental Biology, D-72076 Tübingen, Germany

<sup>4</sup> Institute of Plant Science and Resources, Okayama University, 2-20-1 Chuo, Kurashiki, 710-0046, Japan.

#### **Abstract**

Genes and transposons can exist in variable DNA methylation states with potentially differential transcription. However, how these epialleles emerge is poorly understood. Here we show that crossing a hypomethylated plant with its normally methylated wild-type counterpart results already in the F1 generation in widespread changes in DNA methylation and transcription patterns. Novel non-parental and heritable epialleles arise at many genic loci, including a locus that itself controls DNA methylation patterns, but with most of the changes affecting pericentromeric transposons. While a subset of transposons show immediate resilencing, a large number displays decreased DNA methylation, which is associated with de novo or enhanced transcriptional activation. This can translate into transposon mobilization in the progeny. Our findings reveal that the combination of distinct epigenomes can be viewed as an epigenomic shock, which is characterized by a round of epigenetic variation creating novel patterns of gene and TE regulation.





## Introduction

In eukaryotic genomes, cytosine methylation represents an epigenetic mark involved in the silencing of transposable elements (TEs), genes, and transgenes (Law et al. 2010, Rigal et al. 2011). In plant genomes, TEs are typically silent and associated with dense DNA methylation in the three cytosine contexts CG, CHG and CHH (where H is any base but G). Repression of gene transcription by DNA methylation often correlates with methylation of promoter sequences, whereas transcriptionally active protein-coding genes tend to be methylated exclusively at CG positions in their bodies (Zhang et al. 2006, Zilberman et al. 2007, Cokus et al. 2008, Lister et al. 2008).

The inbreeding plant *Arabidopsis thaliana* (Arabidopsis) represents a unique model species for which the molecular mechanisms involved in maintaining and modulating DNA methylation patterns have been extensively dissected, and for which the genome-wide distribution and stability of DNA methylation in numerous natural accessions and mutants has been established (Becker et al. 2011, Schmitz et al. 2011, Schmitz et al. 2013, Stroud et al. 2013, Hagmann et al. 2015). In Arabidopsis, faithful propagation of CG methylation patterns upon de novo DNA synthesis during DNA replication is safeguarded by the DNA methyltransferase METHYLTRANSFERASE 1 (MET1), the plant homolog of human DNA methyltransferase 1 (DNMT1), such that symmetrical CG sites in the genome are usually fully methylated or not at all (Kankel et al. 2003, Saze et al. 2003). Maintenance of non-CG methylation is more complex and involves the partially redundant activities of the DNA methyltransferases DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), CHROMOMETHYLASE 2 (CMT2) and CHROMOMETHYLASE 3 (CMT3). The three proteins act in self-reinforcing loop mechanisms that also rely on histone H3 methylation at lysine 9 (H3K9me) and mostly small RNAs of 24-nt in length (Du et al. 2012, Zemach et al. 2013, Du et al. 2014, Stroud et al. 2014). The DECREASE IN DNA METHYLATION 1 (DDM1) chromatin remodeler also controls DNA methylation, likely by allowing methyltransferases to access their H1-containing heterochromatic DNA targets (Zemach et al. 2013).

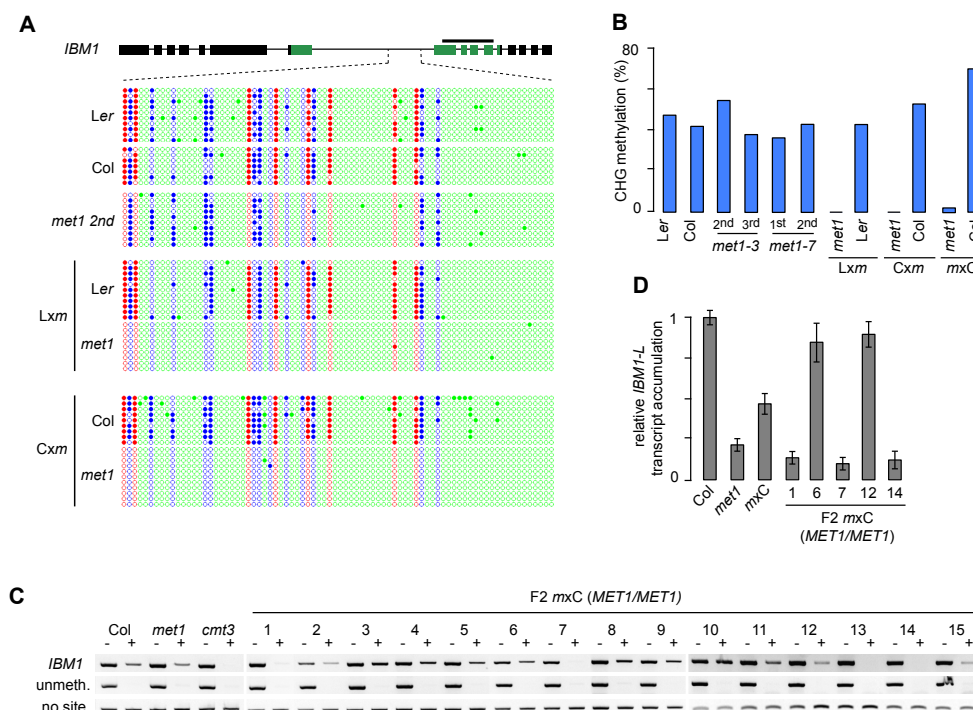
Similar to changes in DNA sequence, differences in DNA methylation, either of natural, spontaneous origin or experimentally-induced, can impact genome stability, gene expression and phenotypic variation. Deficiencies in *DDM1* induce drastic hypomethylation of heterochromatin at all cytosine contexts, resulting in transcriptional upregulation of some heterochromatic genes, a massive activation of TE transcription, and mobilization of various TEs (Vongs et al. 1993, Miura et al. 2001, Singer et al. 2001, Lippman et al. 2003b, Tsukahara et al. 2009, Fu et al. 2013, Zemach et al. 2013).



Elimination of MET1 in null *met1* mutants causes a near-complete loss of CG methylation at genes and TEs, which is associated with misexpression of numerous protein-coding genes and broad transcriptional activation of TEs (Saze et al. 2003, Zhang et al. 2006, Cokus et al. 2008, Lister et al. 2008). However, although transcriptionally activated, some autonomous TEs remain immobile, owing to other silencing mechanisms; transposition of the *Evadé* (*EVD*) retrotransposon occurs only after several generations of inbreeding or in combinations of *met1* with mutations in other silencing factors (Kato et al. 2003a, Mirouze et al. 2009, Rigal et al. 2011).

In addition to the loss of CG methylation, *met1* mutants display substantial ectopic CHG and CHH hypermethylation at thousands of genes and TEs (Mathieu et al. 2007, Lister et al. 2008, Stroud et al. 2013). Depending on the target, this has been linked to erratic/mis-targeting of RNA-directed DNA methylation (RdDM), transcriptional silencing of the gene encoding the REPRESSOR OF SILENCING 1 (ROS1) DNA demethylase, and mis-regulation of the *INCREASE IN BONSAI METHYLATION 1* (*IBM1*) gene, which encodes a H3K9me2 demethylase that specifically targets protein-coding genes (Mathieu et al. 2007, Saze et al. 2008, Rigal et al. 2012). *IBM1* contains a large intron with a heterochromatin-like segment that is methylated at CG and non-CG sites. Lack of either type of methylation in this region disrupts the heterochromatin structure and is associated with impaired production of the long *IBM1* transcript *IBM1-L*, which encodes the functional form of the IBM1 H3K9me2 demethylase (Rigal et al. 2012, Saze et al. 2013b). In addition to the ROS1- and IBM1-mediated defects in *met1* mutants, genes pre-marked with H3K27me3 tend to gain ectopic H3K9me2 and to be depleted in H3K27me3 in *met1*, while heterochromatic loci show depletion in H3K9me2 and ectopic H3K27me3 upon CG methylation loss (Mathieu et al. 2005a, Deleris et al. 2012).

In *ddm1* and *met1* mutants, changes in DNA methylation generate epialleles that remain unaltered in their DNA sequence. In a few examples, such epialleles have been associated with heritable differential transcription (Jacobsen et al. 1997, Soppe et al. 2000b, Stokes et al. 2002, Kankel et al. 2003, Saze et al. 2007b). It has been long established that some *ddm1*- and *met1*-induced hypomethylated epialleles can be stably transmitted through several generations after outcrossing the causing mutation (Vongs et al. 1993, Kakutani et al. 1999). Similarly, hypermethylated epialleles can be transmitted across generations, though they appear to be less stable than hypomethylated ones (Jacobsen et al. 1997, Jacobsen et al. 2000a, Saze et al. 2007b).



**Figure 1.** Creation of a non-parental *IBM1* epiallele in F1 epihybrids.

(A) Bisulfite sequencing analysis of DNA methylation at the large *IBM1* intron in the indicated genotypes. Filled and empty circles indicate methylated and unmethylated cytosines, respectively. The color of the circle denotes cytosine sequence context (red, CG; blue, CHG; green, CHH). (B) CHG methylation level of the large *IBM1* intron in the indicated genotypes as determined by bisulfite sequencing. Parental origin of *IBM1* alleles in Cxm and *mxC* F1 plants was assigned based on presence/absence of CG methylation. (C) DNA methylation analysis at the large *IBM1* intron using *AluI* digestion (inhibited by CHG methylation) followed by PCR. For each genotype, PCR amplification was performed on digested (+) and undigested DNA (-). A region unmethylated in the WT and a region deprived of *AluI* site were amplified as controls of digestion efficiency and DNA loading, respectively. (D) RT-qPCR analysis of the accumulation of *IBM1* full length mRNA (*IBM1-L*) in the indicated genotypes."

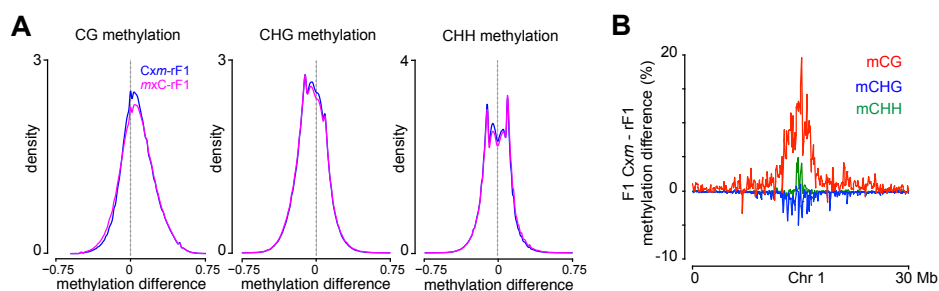
To further analyze transgenerational epigenetic inheritance, segregation of DNA methylation variants (epialleles) was induced in two populations of epigenetic recombinant inbred lines (epiRILs) that were created by outcrossing either a *met1* or a *ddm1* hypomethylated parent plant to an isogenic wild-type plant (Johannes et al. 2009a, Reinders et al. 2009b). Studies of these two epiRIL populations have confirmed overall stable inheritance of parental DNA methylation patterns, but also revealed some degree of epigenetic instability of parental epialleles. In *ddm1*-epiRILs, several TEs and single copy sequences derived from the mutant parent undergo progressive DNA remethylation in a process requiring siRNA biogenesis (Teixeira et al. 2009), while in later generations of some *met1*-derived epiRIL lines non-parental methylation patterns can be detected, revealing the creation of further epigenetic novelty in these plants (Reinders et al. 2009b). Additionally, although *EVD* and *CACTA* TEs are immobile in both parents, proliferation of the two TEs was detected in a fraction of *met1*-derived epiRILs (Mirouze et al. 2009, Reinders et al. 2009b), reinforcing the notion that parental epigenetic regulatory circuits have been altered in these lines.

Here we examined the immediate consequences of combining the contrasting wild-type and *met1* methylomes in F1 plants. Using whole genome DNA methylation analysis in combination with sequencing of mRNA and small RNA transcriptomes, we reveal widespread non-additive variation in DNA methylation, TE silencing and gene expression. These effects translate into the creation of novel heritable epigenetic variants and changes in silencing control and efficiency.

## Results

### Emergence of a novel non-parental heritable hypomethylated *IBM1* epiallele in F1 epihybrids

We previously reported that proper transcription of *IBM1* requires the concomitant presence of CG and CHG methylation in the largest intron of the gene. To determine the role of *MET1* in methylation of *IBM1* intron sequences, we crossed a *met1-3* mutant (hereafter referred to as *met1*) plant in a Col-0 background, which has reduced DNA methylation in this intron, to a Ler-0 *MET1* wild-type (*MET1*<sup>+</sup>) plant. Not only was CG methylation of the *met1*-derived *IBM1* allele not restored in F1 plants (Rigal et al. 2012), Figure 1A), but there was a further, drastic loss of intronic CHG methylation (Figure 1A). This was not due to parental effects of *met1*, since a drastic loss of CHG methylation was not seen upon selfing of *met1* plants over several generations (Figure 1B). This indicated that the drastic loss of CHG methylation was a specific property of interaction between *met1* and *MET1*<sup>+</sup> genomes in F1 hybrid plants.



**Figure 2.** Global DNA methylation changes in F1 epihybrids. (A) Kernel density plots of DNA methylation differences between Cxm or mxG epihybrids and the reconstructed F1 (rF1) at CG, CHG and CHH contexts. (B) Chromosomal view of the mean methylation difference between Cxm and rF1 in 100-kb tiles along chromosome 1 (Chr1). The m prefix denotes methyl.

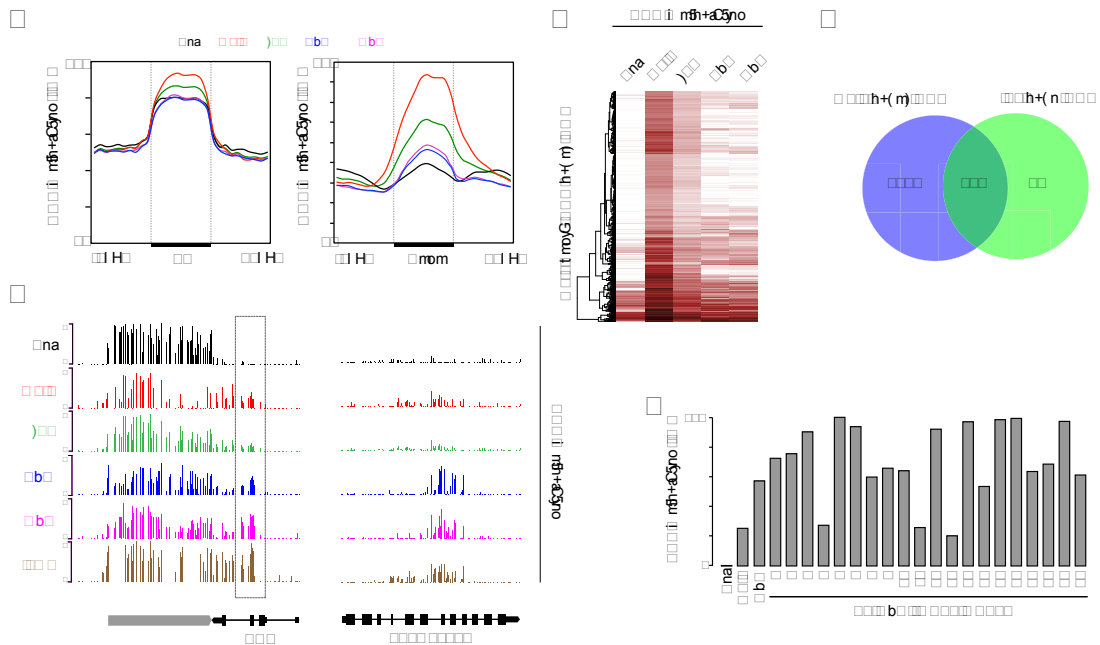
To eliminate effects caused by genetic differences throughout the genome of the Col-0 *met1-3* and the Ler-0 *MET1*<sup>+</sup> parents, we outcrossed a *met1-3* plant to an isogenic Col-0 *MET1*<sup>+</sup> plant. Because the genomes of these parents are nearly identical, but differ drastically in DNA methylation patterns, we named the resulting F1 plants epihybrids. *MET1*<sup>+</sup> plants of the Col-0 and Ler-0 accessions showed similar DNA methylation patterns at the *IBM1* intronic region (Figure 1A and B). Independent of the orientation of the cross (Col-0 x *met1* (Cxm) or *met1* x Col-0 (mxC)), we again observed the emergence of a non-parental DNA methylation pattern at the *met1*-derived allele (Figure 1A and B). Similar results were obtained when we crossed the *met1-7* allele in the Col-0 background to a *MET1*<sup>+</sup> plant of the Van-0 accession (Figure S1A). Together, these results suggest that the appearance of the non-parental hypomethylated *IBM1* epiallele in F1 epihybrids is linked to the difference in DNA methylation patterns between parental plants. The mechanism involved in CHG methylation loss remains to be determined, but it seems to be independent of IBM1 and ROS1 (Figure S2).

To assess heritability of the novel *IBM1* epiallele, we determined *IBM1* intron methylation in F2 progeny of a self-fertilized mxC F1 individual. As expected for Mendelian segregation, one quarter of the *MET1*<sup>+</sup> F2 plants were homozygous for the hypomethylated *IBM1* epiallele; these plants displayed low *IBM1-L* transcript accumulation (Figures 1C and D). Additionally, absence of intronic methylation was inherited in F3 progeny of a self-fertilized *MET1*<sup>+</sup> F2 individual homozygous for the hypomethylated *IBM1* epiallele (Figure S1B). Hence, outcrossing a hypomethylated *met1* parent to a *MET1*<sup>+</sup> plant induces the generation of a novel, non-parental hypomethylated *IBM1* epiallele in epihybrid progeny. The novel epiallele is stably inherited through at least two generations, independently of the presence of the *met1* mutation.

### **DNA methylation variation in F1 epihybrids mainly occurs at pericentromeric TEs**

Our observation at the *IBM1* locus prompted us to examine the extent of DNA methylation variation in epihybrid plants on a genome-wide scale. We performed whole-genome bisulfite sequencing in Col-0 *MET1*<sup>+</sup> and *met1* parents and in both reciprocal F1 epihybrids. We combined the parental datasets to reconstruct a synthetic F1 sample (hereafter named rF1) where each genomic cytosine is methylated at the mid-parent value (MPV). Overall DNA methylation levels were similar in the two reciprocal F1 epihybrids and showed a substantial increase in CG methylation, decreased CHG methylation and localized changes in CHH methylation compared with rF1 (Figure 2A).





moyG *nfh+aDyso* yom (yh+H)y 0  
 2m)Ct m y5yH 5no ne *nfh+aDyso* yo y35m t non5( n0n2m) 0Co t nom0  
 Gho5oyot C5ar05nom *h+(m)* *nfh+aDyso* mC5i C ne *nfh+aDyso*  
 an2m0 k yhyo *h+(m)* *nfh+aDyso* 000nk 0)m( )n0m55hm 0Co 5hm  
 Gha i o0)m( )n0m55hm yo y35m t non5( n0m0moo yC )C )m( )n0m55no ne5hm2m)C  
 Hrk nm t nom0 Gho5oyot C5ar05nom *h+(m)* *nfh+aDyso* yo Co t non0  
 Gho5oyot C5ar05nom *h+(ni)* *nfh+aDyso* yo m(yh+H)y 0 noni mH)nk 0m)  
 2yk 0ne *nfh+aDyso* C55hm yo y35m anG *nfh+aDyso* an2m0C55hm anG 0yo  
 yo y35m t non5( n0k 00 nfh) yom tH+ G y m5no anank m tH+% hmyo y2y - Ca  
 i b (aCo5CoCaTm yoGa m5hmnon0CoCaTm yo y - )m Co

Gain of CG methylation was prominent at pericentromeric regions and peaked at centromeres, where CHG and CHH methylation also showed the strongest deviation from rF1 levels (Figure 2B). Consistently, TEs globally displayed increased CG methylation in epihybrids relative to rF1, whereas CG gene-body methylation was at the expected rF1 level (Figure S3A and S3B). CHG methylation was lower than the rF1 level at both genes and TEs. These results indicate that most of variation from the expected rF1 DNA methylation level occurs at pericentromeric regions of epihybrids.

By comparing the two reciprocal F1 epihybrids to rF1, we determined differentially methylated regions (DMRs) at all three cytosine sequence contexts significantly deviating from the mere 1:1 addition of parental methylomes. We also determined DMRs between the *MET1*<sup>+</sup> and *met1* parents. We defined a total of 9,892 F1 hypomethylated DMRs (hypo-DMRs), where the epihybrids had lost DNA methylation, and 8,892 F1 hypermethylated DMRs (hyper-DMRs), where the epihybrids had gained methylation (Figure S3C). Independent of their type, the vast majority of the identified DMRs matched TE loci, and gains in CG methylation and losses in CHG methylation were the most abundant types of DMRs. In each case, more than 60% of the DMRs were common to reciprocal F1 epihybrids, indicating that most DNA methylation changes in epihybrids occur independently of cross orientation, similar to the creation of the hypomethylated *IBM1* epiallele.

### **Ectopic CHG methylation at genes in F1 epihybrids**

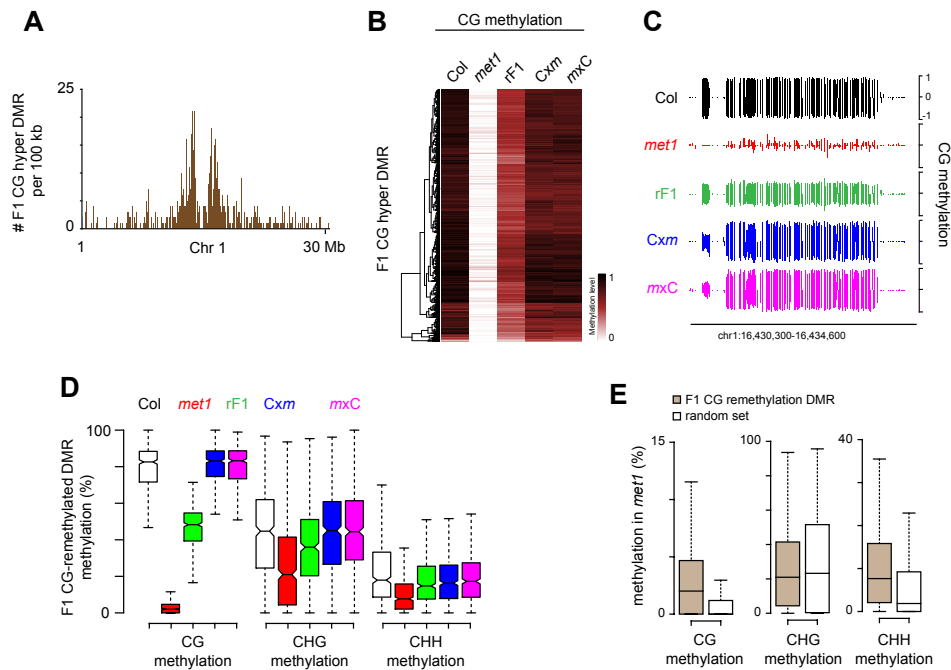
Thousands of genomic loci are ectopically hypermethylated at CHG positions in *met1* mutants, in part due to misregulation of *IBM1* and *ROS1* genes (Mathieu et al. 2007, Lister et al. 2008, Deleris et al. 2012, Rigal et al. 2012). Overall, genes CHG-hypermethylated in *met1*, but not TEs, continued to have a higher CHG methylation level in epihybrids compared to *MET1*<sup>+</sup> plants (Figure 3A, B, C). Using public data (Stroud et al. 2013), we defined genes ectopically methylated at CHG positions in *ibm1* and in the *ros1 dml2 dml3 (rdd)* triple mutant of the DNA demethylases expressed in the sporophyte. Intersecting genes hypermethylated in *met1* and *ibm1* or in *met1* and *rdd* with those hypomethylated in F1 revealed limited overlap, indicating that normal methylation at IBM1- and RDD-targeted genes was mostly not restored (Figure S4A). We previously demonstrated that restoring proper levels of functional *IBM1* transcript accumulation in *met1* resulted in removal of ectopic CHG and H3K9me2 methylation at selected genes (Rigal et al. 2012). Also, we found that *IBM1* and *ROS1* were transcribed at a near mid-parental level in epihybrids (Figures 1C and S4B). Together, these data suggest that the amount of functional IBM1 produced in epihybrids is limiting, thus contributing to persistent ectopic CHG methylation in F1.



Furthermore, about 80% of the genes CHG-hypermethylated in *met1*, and not in either *ibm1* or *rdd*, retained CHG methylation in epihybrids (Figure S4A). Thus, whatever the cause of ectopic genic CHG hypermethylation in *met1*, this methylation largely persists in F1 plants.

We identified about 110 genes with CHG hypermethylation in each epihybrid compared with either parent (Figure S5A). Although at a small subset of genes CHG hypermethylation seemed to be initiated specifically in F1 plants (Figure S5A,B), most transgressive CHG hyper-DMRs already displayed a certain level of ectopic CHG methylation in the *met1* parent (Figure 3D, S5A). About 60% of these genes were also CHG hypermethylated in *ibm1* (Figure 3D and S5C), supporting the notion that limiting IBM1 activity is a major contributor to transgressive CHG hypermethylation in F1 epihybrids. For example, *BONSAI* (*BNS*), the first identified IBM1 target (Saze et al. 2008), gained CHG methylation at the same positions in *ibm1* and, albeit to a lesser extent, in *met1*, and became further hypermethylated in F1 epihybrids (Figure 3D). To determine which *BNS* allele was hypermethylated in epihybrids, we self-pollinated *mx*C F1 plants, and assessed *BNS* methylation state in *MET1*<sup>+</sup> F2 individuals. We did not recover *MET1*<sup>+</sup> F2 plants with a *MET1*<sup>+</sup>-like *BNS* methylation level; about 3/4 of the plants displayed a methylation level that was similar or higher than that of the F1 epihybrid parent, while the other F2 plants showed a *met1*-like level of methylation (Figure 3E). These results suggest that the *MET1*<sup>+</sup>-derived *BNS* allele acquired CHG hypermethylation in F1 epihybrids, creating novel non-parental *BNS* epialleles. Intriguingly, some F2 individuals with high *BNS* CHG methylation levels had normal *IBM1-L* expression (Figure 3E and 1C), suggesting that, unlike the original *BNS* allele in the *MET1*<sup>+</sup> background, the novel CHG hypermethylated *BNS* epiallele formed in epihybrids is largely insensitive to IBM1 activity.

Altogether our data reveal that, similar to *met1*-induced loss of CG methylation, *met1*-induced CHG hypermethylation represents a reservoir of epiallelism, since it is stably inherited at numerous genes in F1 epihybrids. Hypermethylation of CHG sites can also be induced or reinforced in epihybrids and, as exemplified by the *BNS* locus, initiates onto the cognate *MET1*<sup>+</sup>-derived allele, thereby creating novel epialleles upon confrontation of the parental epigenomes in F1 plants.



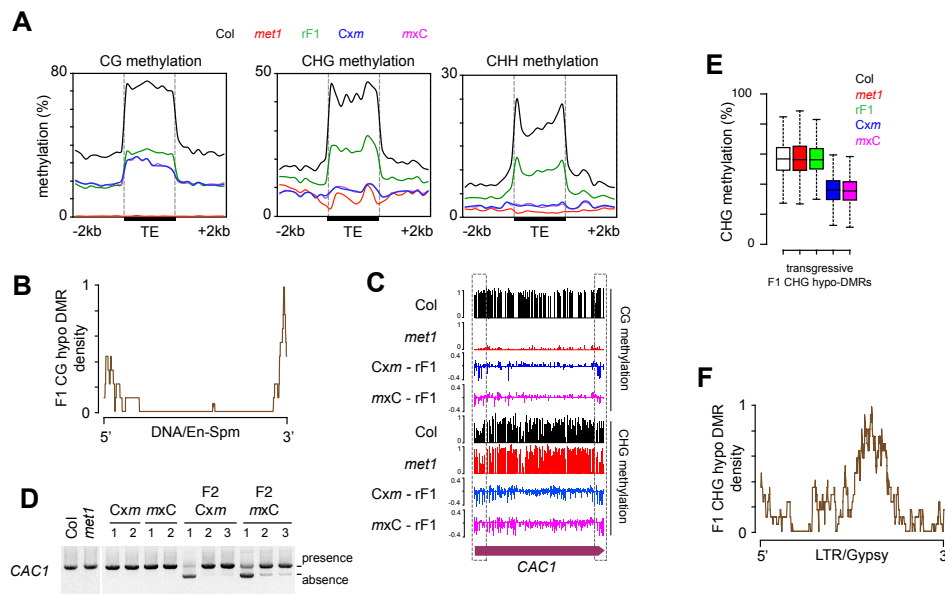
**Figure 4.** CG remethylation in epihybrids.

(A) Chromosomal view of the cumulative number of F1 CG hypermethylation DMRs in 100-kb tiles along chromosome 1. (B) Heat-map of CG methylation levels within F1 CG hypermethylation DMRs. (C) Genome-browser view of CG methylation at a representative genomic region displaying CG remethylation in F1 epihybrids. (D) Box plot of CG, CHG and CHH methylation levels in F1 CG-remethylated DMRs. Values were extracted from one replicate per genotype. (E) Box plot of CG, CHG and CHH methylation levels in F1 CG-remethylated DMRs and in a random set of sequences of equivalent type and length.

### Immediate CG remethylation at TEs in F1 epihybrids

F1 CG hyper-DMRs mostly overlapped with TEs (2,138 TEs, Figure S3C) preferentially located in pericentromeric regions of the chromosomes (Figure 4A). These genomic loci were highly CG methylated in the *MET1*<sup>+</sup> plants and strongly lost CG methylation in *met1* (Figure 4B), indicating that F1 CG-hyper DMRs mostly reflect remethylation of the corresponding *met1*-derived TE epialleles. At about 25% of remethylated loci, CG methylation was fully restored to the *MET1*<sup>+</sup> level in F1 epihybrids (Figure 4C and D). The relative distribution of TEs overlapping with fully remethylated F1 CG hyper-DMRs essentially mirrored the distribution of all TEs in the *Arabidopsis* genome (Figure S6A), showing that CG remethylation in epihybrids affects all types of pericentromeric TEs. Noticeably, compared with an equivalent set of random TE regions, CG-remethylated loci retained more CHH methylation and a substantial CG methylation level in the null *met1-3* mutant (Figure 4E), indicating that part of CG methylation at these TE sequences is independent of MET1 activity. Full restoration of CG methylation in F1 epihybrids was globally accompanied by re-establishment of CHG methylation, whereas CHH methylation did not significantly deviate from the rF1 level (Figure 4D). This suggests that remethylation does not involve DRM2-mediated methylation, which is expected to equally target all cytosine positions for DNA methylation.

We also observed CHH hyper-DMRs in F1 epihybrids, although these were less abundant than CG hyper-DMRs (Figure S3C). CHH hyper-DMRs were overrepresented for pericentromeric LTR/Gypsy retroelements and were associated with DNA hypermethylation in all cytosine contexts, with transgressive CHH and CHG methylation where the F1 had higher methylation than either parent (Figure S6A, B, C). Because of the association specifically with non-CG methylation, this hypermethylation likely occurs through the canonical DRM2-mediated RdDM pathway. In support of this notion, CHH hyper-DMRs showed a higher density at the edges of LTR/Gypsy elements (Figure S6D), which represent preferential target sites for DRM2-mediated methylation (Zemach et al. 2013). Our results thus reveal that numerous LTR/Gypsy elements acquire non-parental CHH methylation patterns at their edges in F1 epihybrids.



**Figure 5.** Hypomethylation at TEs in epihybrids.

(A) Average distribution of CG, CHG and CHH methylation in indicated genotypes over VANDAL elements containing at least one F1 CHH and CHG hypo-DMRs. (B) Distribution of F1 CG hypomethylation DMR density over DNA/En-Spm elements. (C) Genome-browser view of CG and CHG methylation at *CAC1*. Col and *met1* parental methylation patterns are shown together with the *rF1* profile subtracted from the *Cxm* or the *mxC* F1 profiles. The dashed boxes show regions with CG and CHG hypomethylation in epihybrids. (D) PCR assay of *CAC1* excision in indicated genotypes. Amplification of the “absence” band reflects mobilization of the TE. (E) Box plot of CHG methylation levels in transgressive F1 CHG hypomethylation DMRs in indicated genotypes. Values were extracted from one replicate per genotype. (F) Distribution of F1 CHG hypomethylation DMR density over LTR/Gypsy elements.

### Trans-demethylation at TEs in F1 epihybrids

Hypomethylation in F1 epihybrids primarily affected TEs and was mostly due to changes at CHG and CHH positions (Figure S3C). F1 CHH hypo-DMRs often displayed concurrent CHG hypomethylation (Figure S7A), suggesting that CHH and CHG hypomethylation occur at least partly through a common mechanism. TEs overlapping with both F1 CHH and CHG hypo-DMRs were overrepresented for *VANDAL* transposons (Figure S7B,C), which were characterized by very low DNA methylation in *met1* mutants (Figure 5A), indicating that DNA hypomethylation in epihybrids affects TE copies inherited from the *MET1*<sup>+</sup> parent.

A specific *VANDAL21* element, *Hi*, transposes in *ddm1* and *met1* mutant backgrounds (Tsukahara et al. 2009, Fu et al. 2013). The new *Hi* copies induce trans-demethylation, mediated by the *Hi*-encoded VANC protein, with a high specificity towards *VANDAL21* family elements and being more pronounced at non-CG sites (Fu et al. 2013). In our material, *Hi* was hypomethylated in *met1* at all cytosine sequence contexts and associated with strong non-CG hypomethylation in F1 epihybrids (Figure S8A). Other *VANDAL21* family members also displayed non-CG hypomethylation in F1 epihybrids (Figure S7C). We detected new *Hi* copies in *met1* and in F1 plants (Figure S8B), supporting that transposed *Hi* copies may induce trans-demethylation of *VANDAL21* elements in F1 epihybrids. Because many other elements from other *VANDAL* families were also strongly hypomethylated in F1 epihybrids (Figure S7C), we sought to determine whether the same type of transposition-dependent trans-demethylation mechanism was engaged in these families. We determined TE copy number genome-wide by comparing the methyl-C DNA sequencing coverage of *met1* and F1 epihybrid plants with that of the *MET1*<sup>+</sup> parents. Validating this approach, we detected an increase in *EVD* and *Hi* copy number in the second generation of the *met1* homozygous line (Figure S8C). Additionally, we detected more DNA copies of the *ATCOPIA21* element *AT5TE65370* in *met1*, which we named *Complice*, previously found to be active only in advanced generations of *ddm1* mutants (Tsukahara et al. 2009); Figure S8C). *Hi* copy number further increased in *Cxm* but not in the reciprocal *mxC* F1 epihybrids (Figure S8B,C), suggesting that the male gametophyte represents a developmental window for proliferation of activated *Hi* elements. Alternatively, *MET1* activity may be required in the male gametophyte to restrain *Hi* transposition. Importantly, this analysis did not reveal mobilization of other *VANDAL* elements in either *met1* or F1 epihybrids, suggesting that loss of DNA methylation occurring at *MET1*<sup>+</sup>-derived *VANDAL* elements in F1 epihybrids is neither necessarily caused by nor associated with transposition of TE members from the corresponding hypomethylated families.





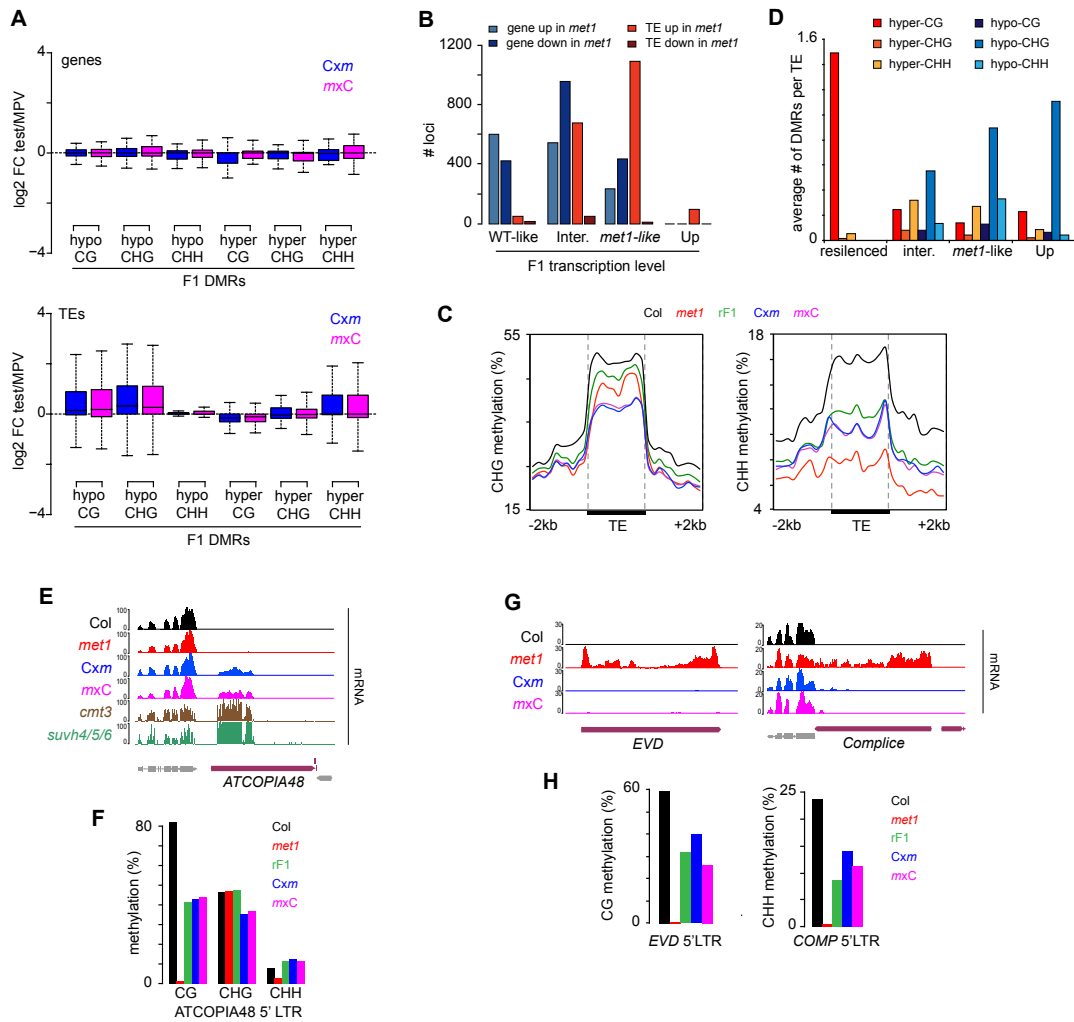
Considering the previously reported high specificity for *Hi*-induced trans-demethylation, we conclude that production of the VANC anti-silencing factor from transcriptionally activated *VANDAL* elements in *met1* is sufficient to induce trans-demethylation of elements of the same family in F1 without requirement for transposition.

### **Trans-demethylation is associated with TE mobilization**

The CG hypomethylation of *VANDAL* elements was more pronounced in their 3' half, which contains the ORF encoding VANC (Figure 5A). F1 CG hypo-DMRs also overlapped with DNA/En-Spm DNA transposons (Figure S7B), and CG hypomethylation was preferentially located near the borders of these elements (Figure 5B). Studies of the maize Spm transposon have revealed that the transposon-encoded TnpA protein can induce DNA hypomethylation at TnpA binding sites that are located at the element ends, and promote its transcription (Schlappi et al. 1994, Cui et al. 2002). The distribution of F1 CG hypo-DMRs suggests that a similar mechanism may operate at *ATENSPM* elements in epihybrids. The En-Spm elements *CACTA1* (*CAC1*) and *CAC2* transpose in *ddm1* mutants and *ddm1*-epiRILs (Miura et al. 2001, Johannes et al. 2009a, Tsukahara et al. 2009). *CACTA* movement has also been detected in *met1*-epiRILs, although the mobilized elements were not assigned to a specific element (Reinders et al. 2009b). *CAC1* is mobilized in *met1 cmt3* but not in *met1*, indicating that both CG and CHG methylation cooperate to repress *CAC1* activity (Kato et al. 2003a). In F1 epihybrids, *CAC1* edges were hypomethylated at CG sites and also displayed transgressive CHG hypomethylation (Figure 5C). Because CG methylation was completely lost in the *met1* parent, this indicates hypomethylation of the *MET1*<sup>+</sup>-derived *CAC1* allele in the epihybrids. The hypomethylation of *CAC1* at both CG and CHG sites in F1 epihybrids correlated with mobilization of this element in F2 progeny (Figure 5D). Together, these data show that DNA hypomethylation induced in F1 epihybrids, potentially involving the transposon-encoded TnpA protein, results in mobilization of an element that did not transpose in either parent.

### **Transgressive hypomethylation of LTR/Gypsy elements in F1 epihybrids**

Transgressive CHG hypomethylation was specifically initiated in F1 epihybrids at one thousand genomic regions, which were highly methylated in both *MET1*<sup>+</sup> and *met1* parents (Figure 5E and S9A) and were overrepresented for LTR/Gypsy elements of long size (mean length of 6.43 kb; Figure S9B and S9C). Transgressive CHG hypomethylation preferentially occurred in internal coding regions of LTR/Gypsy TEs, with a bias towards the 3' half of the elements (Figure 5F), and it was accompanied by decreased CHH methylation within TE bodies, with little change at TE edges (Figure S9D).



**Figure 6.** Transcription in F1 epihybrids

(A) Box plots of the log<sub>2</sub> ratio of RPKM values of genes (top panel) and TEs (bottom panel) overlapping with indicated DMRs in epihybrids over the mid-parent value (MPV). (B) Number of genes and TEs mis-regulated in *met1* grouped according to their transcription level in epihybrids (Inter., intermediate). Only genes showing equivalent transcriptional destiny in both epihybrids are shown. (C) Average distribution of CHG and CHH methylation in indicated genotypes over TEs activated in *met1* and transcribed at a *met1*-like or higher level in F1 epihybrids. (D) Average number of indicated DMRs overlapping with TEs grouped according to their transcription level in epihybrids (Inter., intermediate). (E) Genome browser view of mRNA accumulation at *ATCOPIA48* (*AT1TE66380*) in indicated genotypes. RNA levels in *cmt3* and *suvh4/5/6* are from (Stroud et al, 2012). (F) DNA methylation levels at the 5' LTR of *ATCOPIA48*. (G) Genome browser view of mRNA accumulation at *EVD* (*AT5TE20395*) and *Complice* (*AT5TE65370*) elements in indicated genotypes. (H) DNA methylation levels at the 5' LTRs of *EVD* and *Complice* elements."

CMT2 targets the body of long TEs for CHH methylation, and to a lesser extent for CHG methylation (Zemach et al. 2013, Stroud et al. 2014). *CMT2* transcripts were reduced three-fold in *met1*, potentially contributing to decreased non-CG methylation at long TEs, while *CMT2* transcripts accumulated at a mid-parental level in F1 plants (Figure S9E). Reduced *CMT2* expression relative to *MET1*<sup>+</sup> plants may be at the origin of CHH hypomethylation at internal sequences of long TEs in F1 epihybrids. Collectively, our results reveal widespread DNA hypomethylation at both DNA transposons and retroelements in F1 epihybrids.

### Transcription dynamics in F1 epihybrids

It is generally accepted that DNA methylation can influence transcription and transcript accumulation, although the mechanisms are often unclear. Because of the profound modification of DNA methylation patterns occurring in F1 epihybrids, we determined how this correlated with genome-wide mRNA profiles of F1 epihybrids and their parents. At the chromosomal level, pericentromeric regions retained high transcript accumulation in epihybrids relative to *MET1*<sup>+</sup> parents (Figure S10A), extending previous reports of inheritance of *met1*-induced release of silencing at selected loci (Kankel et al. 2003, Lippman et al. 2003b, Reinders et al. 2009b). Additional changes in transcript profiles indicated that the silencing/desilencing state of numerous loci is not stably inherited from *met1* parents by the F1. Overall gene transcript levels varied only modestly between parents and F1 epihybrids, independent of DNA methylation changes in F1 (Figure 6A and S10B). In contrast, TEs that were strongly activated in *met1* were expressed at levels above the MPV in F1 epihybrids, and both TEs overlapping with F1 CG and CHG hypo-DMRs and F1 CHH-hyper DMRs showed increased expression relative to the MPV (Figure 6A and S10B). This indicates that changes in DNA methylation in epihybrids are mostly associated with further derepression of TE silencing.

Combining two complementary approaches to determine differential expression between samples, we identified 5,775 misregulated loci in *met1* compared with the *MET1*<sup>+</sup> parent (3,606 upregulated, 2,169 downregulated) (Figure S10C), substantially adding to the number of *met1*-misexpressed loci previously identified (Zhang et al. 2006, Zilberman et al. 2007, Lister et al. 2008, Dowen et al. 2012a). Next, we determined how loci that were mis-expressed in the *met1* parents were expressed in F1 epihybrids (Figure 6B). The large majority (more than 85%) of loci mis-regulated in *met1* behaved in a similar manner in both F1 epihybrids. Of the 1,576 genes up-regulated in *met1*, only about 40% accumulated at an intermediate level in epihybrids, indicating stable inheritance of *met1*-induced transcriptional derepression in epihybrids.



This set of genes includes and behaves similarly to the archetype *FWA* gene that remains transcriptionally derepressed and hypomethylated upon restoration of the MET1 function (Kankel et al. 2003). Expression of another 40% of *met1*-derepressed genes was corrected in epihybrids, returning to *MET1*<sup>+</sup> levels (Figure 6B), revealing efficient reestablishment of transcriptional repression of the corresponding *met1*-derived alleles. The remaining genes, about 240, were further upregulated in F1 and were expressed at *met1*-like levels. Independent of their behavior in F1, genes up-regulated in *met1* showed no remethylation in F1 plants (Figure S11A), indicating that restoration of normal expression levels often does not depend on DNA methylation. There were two notable exceptions: *SDC*, a well known target of DNA methylation dependent silencing (Henderson et al. 2008), and *KELP (AT4G10920)*, which encodes a target of tobamoviruses (Matsushita et al. 2001), were efficiently re-silenced in F1 epihybrids. Re-silencing correlated with DNA remethylation of tandem repeats located in the promoter regions of both genes (Figure S12).

Similar to what we had observed for *IBM1* and *ROS1*, expression of most *met1*-downregulated genes was not restored to *MET1*<sup>+</sup>-levels in F1 epihybrids (Figure 6B). Thus, similar to transcriptional activation, *met1*-induced transcriptional repression can be inherited. Transcriptional down-regulation in *met1* was not associated with hypermethylation and the F1 methylation pattern at these genes was similar to the expected rF1, independent of their F1 expression level (Figure S11B). This indicates that transcriptional changes at these genes in epihybrids mostly result from indirect effects rather than alterations of their DNA methylation patterns.

Transcriptional mis-regulation is not inherited in epihybrids at about 40% of the genes derepressed in *met1*. At these genes, re-establishment of WT transcription is independent of remethylation, indicating that the mechanism preventing formation of stable epialleles at these genes does not rely on DNA methylation. For the other *met1* activated genes and for most *met1* repressed genes, the mutant transcriptional state is stably inherited or exacerbated in F1 epihybrids. Reexamining published data (Zilberman et al. 2008), we found that only genes that were upregulated in *met1* but express in F1 at an intermediate or *met1*-like level were enriched in H2A.Z in *met1* (Figure S13), even though all genes equally lost CG methylation in *met1* (Figure S11). This suggests that H2A.Z enrichment caused by CG methylation depletion (Zilberman et al. 2008, Coleman-Derr et al. 2012a) possibly plays a role in stabilizing transcriptional derepression upon restoration of MET1 function in epihybrid plants. Additionally, as H2A.Z enrichment correlates with gene responsiveness (Coleman-Derr et al. 2012a), these transcriptionally activated epialleles may acquire new sensitivity to developmental and environmental stimuli.



### Further release of TE silencing in F1 epihybrids

In sharp contrast with genes, very few TEs derepressed in *met1* were resilenced in epihybrids; out of the 1,921 TEs activated in *met1*, only 55 showed resilencing (Figure 6B). Unexpectedly, more than half of *met1*-activated TEs accumulated to *met1*-like or even higher levels in epihybrids, while only a third was expressed at MPV, reflecting stable inheritance of their altered transcriptional state in F1 (Figure 6B). For instance, *Hi* and *CAC1*, which were hypomethylated at the *MET1*<sup>+</sup>-derived copies in F1 plants (Figure 5C and S8A), were expressed at *met1*-like levels in epihybrids (Figure S14A), suggesting release of silencing at the corresponding *MET1*<sup>+</sup>-derived alleles in the F1.

Overall, TEs hyperactivated in epihybrids were enriched for LTR retroelements, in particular pericentromeric LTR/Gypsy *ATHILA* retroelements (Figure S14B), which showed CHH hypomethylation in their body and strong transgressive CHG hypomethylation (Figure 6C,D). Transcript accumulation in epihybrids was especially strong for the 3' half of LTR/Gypsy elements (Figure S14C), a region known as *Transcriptionally Silent Information (TSI)* at some *ATHILA* elements (Steimer et al. 2000), which also displayed preferential transgressive CHG hypomethylation in F1 (Figure 5F). Besides LTR/Gypsy retroelements, several LTR/Copia retrotransposons, including *ATCOPIA48* (*ATITE66380*), were transgressively upregulated in F1 plants (Figure 6E and S14B). *ATCOPIA48* LTRs are rich in CG and CHG sites (11 and 17 sites, respectively), which are methylated at average levels of 82% and 46% in *MET1*<sup>+</sup> (Figure 6F). *ATCOPIA48* was expressed at a very low level in *met1* despite loss of CG methylation, indicating that release of silencing in F1 epihybrids likely reflects alteration in an additional repressive mark. CHG methylation was reduced at the *ATCOPIA48* 5' LTR in F1 epihybrids (Figure 6F). Using published data (Stroud et al. 2012), we found that *ATCOPIA48* is strongly activated in *cmt3* and *svh4/5/6* mutants (Figure 6E), revealing that its silencing depends on CHG methylation. Therefore, derepression of *ATCOPIA48* in F1 epihybrids likely results from the transgressive CHG hypomethylation.

Collectively, our results show that silencing is further released at numerous TEs in epihybrid plants in association with transgressive CHG hypomethylation.

### Immediate resilencing at TEs in F1 epihybrids

The few TEs with *MET1*<sup>+</sup>-like expression in epihybrids were enriched for DMRs CG-hypermethylated in relative to the rF1 (Figure 6D), suggesting that resilencing at these elements relies on CG remethylation.



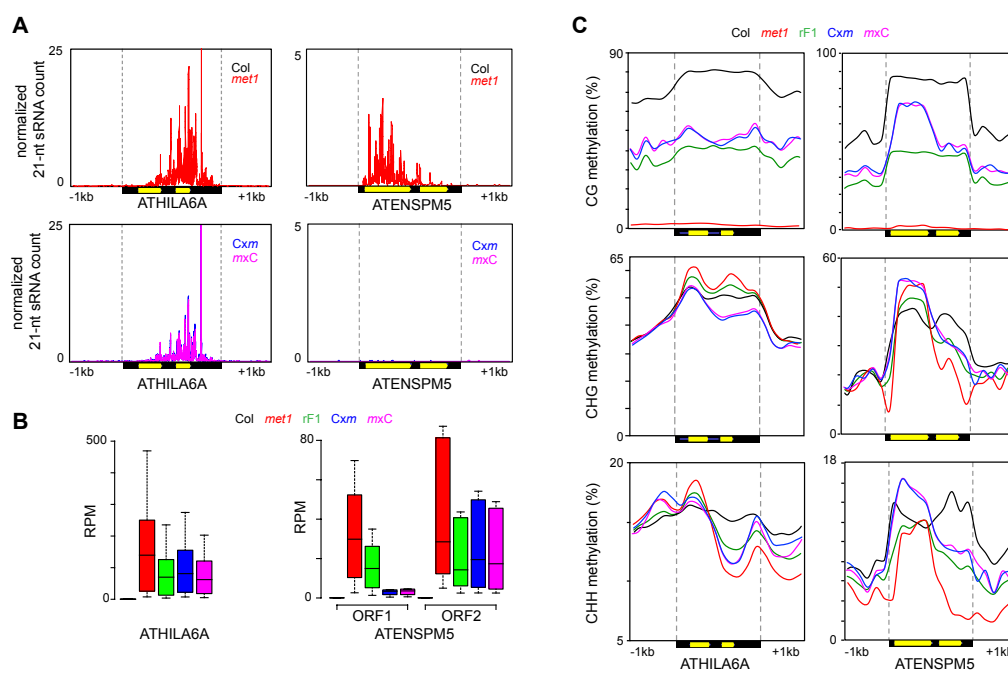


Silencing of the *EVD* and *Complice* retroelements, which are expressed and mobilized in *met1*, was efficiently reestablished in F1 epihybrids (Figure 6G and S8C). *EVD* 5' LTR, which is almost exclusively methylated at CG sites (Mirouze et al. 2009), was not remethylated in F1 plants (Figure 6H). *Complice* LTRs, which lack CG and CHG sites, have around 25% CHH methylation in *MET1*<sup>+</sup>, which is nearly lost in *met1* mutants. Similar to *EVD*, *Complice* 5' LTRs were not remethylated in either reciprocal epihybrids (Figure 6H). Thus, although the molecular mechanism remains to be determined, these results reveal that resilencing of active autonomous TEs originally repressed by DNA methylation, and of their neo-copies, can occur rapidly and independently of DNA remethylation.

### **DNA methylation changes in F1 are associated with concurrent changes in 24-nt sRNA accumulation**

In plants, small RNAs are involved in the establishment of DNA methylation and part of its maintenance in asymmetric context. We deep-sequenced small RNAs from both parents and F1 epihybrids from mature leaves, the same tissue used to profile both mRNA accumulation and DNA methylation, and reconstructed an rF1 sample from the union of the parental datasets. We defined regions of differential 24-nt sRNA accumulation between parents and followed their inheritance by comparing epihybrids with rF1. In *met1*, we identified about twice as many regions of 24-nt sRNA over-accumulation (13,871) than regions of decreased 24-nt sRNA abundance (6,582) (Figure S15A). However, loss of 24-nt sRNA showed much higher amplitude than the increase (99 vs. 13 fold, respectively), accounting for the global reduction in 24-nt sRNAs in *met1* (Figure S15B). Although parental 24-nt sRNAs levels were overall inherited in F1 epihybrids, they were further reduced or increased at a subset of regions enriched in MuDR and LTR/Gypsy TEs, respectively (Figure S15A and C).

We next determined whether DNA methylation changes in epihybrids correlated with local variation in 24-nt sRNA accumulation (Figure S16A, B). F1 CG hypermethylation-DMRs had reduced 24-nt sRNA levels in *met1*, but were close to *MET1*<sup>+</sup> or even higher in the F1. This indicates that CG remethylation in epihybrids likely correlates with 24-nt sRNA production, which requires CG methylation at these loci. F1 non-CG hypermethylation-DMRs, which mostly overlap ends of LTR/Gypsy TEs, showed high transgressive 24-nt sRNA production in F1 epihybrids (Figure S16A, C). Conversely, F1 hypomethylated-DMRs, accumulated low amounts of 24-nt sRNAs in *met1* as well as in epihybrids, suggesting that 24-nt sRNA production is impaired also at the corresponding *MET1*<sup>+</sup>-derived alleles in F1 epihybrids (Figure S17B). For instance, VANDAL elements hypomethylated in epihybrids generated drastically reduced levels of 24-nt sRNA in epihybrid plants (Figure S16D).



**Figure 7.** 21-nt sRNA accumulation, mRNA and DNA methylation levels at ATHILA6A and ATENSPM5 elements. (A) Accumulation of 21-nt sRNAs along ATHILA6A and ATENSPM5 elements in parent (top) and F1 epihybrids (bottom). The two open reading frames in both TEs are depicted as yellow arrows. (B) Box plots of mRNA accumulation at ATHILA6A (left) and the two ATENSPM5 ORFs (right) in indicated genotypes. (C) Average distribution of CG, CHG and CHH methylation in indicated genotypes over ATHILA6A and ATENSPM5 elements producing 21-nt sRNAs in *met1*.

Together, these data show that DNA methylation changes at TEs in F1 epihybrids are largely associated with changes in 24-nt local sRNA accumulation.

In contrast, CHG hypermethylation at genes in epihybrids, including IBM1-targets such as *BNS* and *AT3G20790*, was not associated with accumulation of sRNAs in either parent plant or in epihybrids (Figure S17). This suggests that hypermethylation at these genes in epihybrids is caused by impaired IBM1 activity and/or H3K9me2 enrichment rather than by a sRNA-mediated process, although sRNAs might be transitorily produced and required to target DNA methylation during an earlier developmental window. Similarly, *ddm1*-induced *BNS* DNA hypermethylation depends on H3K9me2 and not on sRNAs (Sasaki et al. 2012).

### **Differential inheritance of *met1*-induced easiRNA production in F1 epihybrids**

When transcriptionally activated, a number of TEs generate 21-nt sRNAs called easiRNAs, for epigenetically activated small RNAs (Lister et al. 2008, Tanurdzic et al. 2008, Slotkin et al. 2009, Creasey et al. 2014). *met1* leaves, similar to flower buds (Lister et al. 2008), were characterized by a global over-accumulation of 21-nt siRNAs matching TEs, suggesting that production of easiRNAs also takes place in the *met1* sporophyte (Figure S15B, S18A). Regions of increased abundance of 21-nt sRNA in *met1* leaves occurred mostly at TEs belonging to the LTR/Gypsy and En-Spm superfamilies (Figure S19B). Few regions of differential 21-nt sRNA accumulation were identified in epihybrids compared with rF1 (Figure S18A), suggesting that easiRNA production from *met1*-derived TE alleles is generally inherited in F1 plants. However, out of the 42 En-Spm elements producing easiRNA in *met1*, 32 consistently showed reduced easiRNA accumulation in reciprocal epihybrids (half of these were *ATENSPM5* elements) and a small subset of 14 *ATHILA* TEs showed further increased accumulation of 21 sRNAs in epihybrids (Figure S18A and B).

Similar to previous results from *ddm1* inflorescences (Nuthikattu et al. 2013), the *ATHILA6A* and *ATHILA2* LTR-retrotransposon families contributed the highest amount of easiRNAs in *met1* leaves, while *ATENSPM5* represented the first contributing DNA transposon family (Figure S18C). Accumulation of easiRNAs was biased towards the 3' *TSI* part of *ATHILA6A* elements, and towards the transposase-encoding ORF1 of *ATENSPM5* elements (Figure 7A). Production of easiRNAs from these TEs was, however, radically different in F1 epihybrids. While the overall F1 level of *ATHILA6A* easiRNAs was similar to rF1, *ATENSPM5* elements accumulated virtually no easiRNAs in epihybrids (Figure 7A). The loss of easiRNA was associated with transcriptional re-silencing of the *ATENSPM5* ORF1 in F1 epihybrids, while the second ORF that produced no or very few easiRNA in *met1*, remained transcriptionally active (Figure 7B).



At *ATHILA6A* elements, the 3' *TSI* region was strongly derepressed in *met1* and remained transcribed at a mid-parent level in epihybrids (Figure 7B). This shows that the accumulation of easiRNAs from *ATHILA6A* and *ATENSPM5* in parent plants and F1 epihybrids parallels the transcriptional activity of their cognate genomic regions. EasiRNAs have previously been implicated in re-establishment of silencing at TEs following *ddm1*-induced transcriptional activation and restoration of functional DDM1 protein by outcrossing (Nuthikattu et al. 2013). Our results indicate that the capacity to function in corrective reestablishment of silencing is not a general feature of all 21-nt easiRNAs.

Re-silencing of *ATENSPM5* ORF1 in F1 epihybrids correlated with DNA hypermethylation at all cytosine contexts and increased 24-nt sRNA accumulation (Figure 7C and S18D). *ATHILA6A* elements, which also accumulated high levels of 24-nt sRNA in *met1* and F1 epihybrids, were rather hypomethylated at CHG sites in epihybrids (Figure 7C and S18D). Together, our results show that easiRNA and/or 24-nt sRNAs produced from different TEs activated in *met1* have distinct consequences on DNA methylation and associated silencing in F1 epihybrids. Interestingly, *ATHILA* elements, which are not resiled in F1, are non-autonomous retrotransposons, whereas most easiRNA-producing *ATENSPM5* elements are long intact DNA TEs.

## Discussion

We have shown that confronting the highly divergent epigenomes of *met1* and Col-0 *MET1*<sup>+</sup> plants, which contain nearly identical genomes, leads to an epigenomic shock characterized by immediate deviation from the expected union of parental methylomes and transcriptomes. Our data show that most of the deviation from mid-parental expectations in F1 epihybrids takes place independently of parental cross orientation, and is therefore independent of differential *MET1*<sup>+</sup> and *met1* gametophyte contribution. This epigenomic shock results in the creation of novel non-parental gene epialleles with modified patterns of non-CG methylation and transcription. Notably, we show that a novel non-parental *IBM1* hypomethylated epiallele with very low transcriptional activity is formed in epihybrids and inherited in subsequent generations. Because *IBM1* protects several thousand genes from H3K9me2 and associated CHG methylation, inheritance of the novel *IBM1* epiallele likely contributes to the appearance of non-parental methylation patterns that have been observed in *met1*-epiRIL lines, the WT progeny of self-pollinated F1 epihybrids (Reinders et al. 2009b).

Most of the changes in DNA methylation occur at TEs in F1 epihybrids. Noticeably, many pericentromeric elements, among which many Helitrons (Figure S6A), exhibit immediate CG remethylation in epihybrids.



At these TEs, CG methylation tightly depends on small RNAs as these loci retain a significant level of CG methylation associated with high amounts of 24-nt sRNAs in the null *met1-3* mutant. However, not all sequences associated with high production of 24 sRNAs in either parent show immediate CG remethylation in epihybrids, suggesting that remethylated loci exhibit specific DNA sequence and/or chromatin features. Our results suggest that CG remethylation in F1 epihybrids is likely independent of the canonical DRM2 *de novo* methylation activity, and we propose that it occurs via sRNA-directed MET1 *de novo* activity (Aufsatz et al. 2004b), which is restored in F1 epihybrids. This activity may be attracted by remaining CG methylation at *met1*-derived alleles and/or by specific sequence or chromatin features. In any respect, immediate CG remethylation in epihybrids prevents establishment of stable hypomethylated epialleles at these loci. Previous analyses of *met1*-derived epiRILs in the F9 generation using the methylation sensitive *HpaII* restriction enzyme have suggested active remethylation of centromeric DNA at CG and/or CHG sites (Reinders et al. 2009b). Our results support this conclusion and further reveal that pericentromeric regions are efficiently targeted for rapid CG remethylation as early as in the F1 generation when MET1 activity is restored. In contrast, DNA remethylation was not detected in the F1 progeny of *ddm1* x Col-0 crosses, but appeared only progressively at TEs located throughout the genome in advanced generations of *ddm1*-epiRILs (Teixeira et al. 2009). This highlights that remethylation occurring upon outcrossing *met1* or *ddm1* mutations involves at least partly distinct molecular processes. Interestingly, our results show that parental CG methylation patterns within gene bodies are overall faithfully inherited in epihybrids (Figure S3B). This suggests the absence of corrective mechanism and/or signal allowing reestablishment of gene-body CG methylation once it has been lost. Alternatively, gene-body remethylation may require multiple generations to proceed following restoration of the WT methylation machinery. Determining genome-wide DNA methylation patterns of late generation *met1*-epiRILs should allow discriminating between these alternatives.

Despite DNA remethylation of heterochromatin, only few *met1*-activated TEs are resilenced in F1 epihybrids. Moreover, we found that two LTR retrotransposons actively transposing in the *met1* parent, namely *EVD* and *Complice*, are transcriptionally resilenced in leaves of F1 epihybrid plants independently of DNA remethylation, revealing the establishment of novel silencing states in F1 epihybrids. Whether *EVD* and *Complice* both acquire similar silencing states, as well as the underlying molecular determinants of these states, remains to be determined. Whatever molecular mechanism involves, restoration of *EVD* silencing in epihybrids is surprising given that *EVD* is mobilized in about 50% of late





generation *met1* epiRILs (Mirouze et al. 2009), and suggests that the novel, DNA methylation-independent, silencing state established at *EVD* in epihybrids can be easily reversed.

Numerous TEs display DNA hypomethylation associated with enhanced release of silencing or de novo transcriptional activation in F1 epihybrids. At some TEs, including elements of the *VANDAL* and *CACTA* superfamilies, hypomethylation likely occurs through trans-demethylation of *MET1*<sup>+</sup>-derived TE alleles mediated by transposon-encoded factors produced by the transcriptionally active *met1*-inherited TE epialleles. We also detected transgressive CHG hypomethylation in epihybrids at numerous TEs, among which a high number of long LTR/Gypsy *ATHILA* elements, which are highly CHG methylated in both *MET1*<sup>+</sup> and *met1* parents. It was previously shown that, only when transcriptionally activated, TEs can be targeted by IBM1 for H3K9me2 demethylation, resulting in CHG hypomethylation and further transcriptional activation (Inagaki et al. 2010b). Accordingly, we previously showed that forcing IBM1 production in *met1* leads to transcriptional over-activation of some TEs (Rigal et al. 2012). We propose that TEs with transgressive CHG hypomethylation are targeted by IBM1 in epihybrids. Because *IBM1* transcription is defective in *met1* (Rigal et al. 2012), these TEs retain H3K9me2 and CHG methylation in the mutant parent despite transcriptional reactivation. In F1 epihybrids, *met1*-derived transcriptionally active TEs may preferentially recruit the limited amount of functional IBM1 produced, resulting in H3K9me2 and CHG demethylation and to further transcriptional derepression. Supporting this view, TEs transcribed at *met1*-like or higher levels in epihybrids, which are overrepresented for *ATHILA* elements, contain a higher number of CHG sites per length of annotation compared to all TEs (Figure S14D), a feature that correlates with sensitivity to *ibm1* mutations (Inagaki et al. 2010b). Internal sequences of long TEs also show CHH hypomethylation in epihybrids, which may originate from reduced *CMT2* transcription in these plants. Further, because *CMT2* mediates CHH methylation through binding to H3K9me2 (Stroud et al. 2014), *CMT2* recruitment at long TEs may be impaired due to decreased H3K9me2/CHG methylation in epihybrids. Because CHH methylation is strongly required for H3K9me2 at *CMT2*-target sites (Stroud et al. 2014), defective *CMT2* recruitment may also contribute to the transgressive reduction in CHG methylation at long TEs in epihybrids. Our data support a scenario in which transgressive CHG hypomethylation and transcriptional upregulation of *ATHILA* in epihybrids occur at *met1*-derived, transcriptionally active, TE copies. Interestingly, a previous study reported transcriptional activation of *ATHILA* in F1 hybrids of the interspecies cross between *Arabidopsis thaliana* and *Arabidopsis arenosa* (Josefsson et al. 2006a).



Supporting our hypothesis, *ATHILA* elements showed basal transcription in *A. arenosa* but not in *A. thaliana*, and upregulation in F1 hybrids specifically affected *ATHILA* elements contributed by the *A. arenosa* parent (Josefsson et al. 2006a). Although *ATHILA* complements of *A. thaliana* and *A. arenosa* genetically differ in terms of DNA sequence and copy number (Josefsson et al. 2006a), it would be interesting to determine whether upregulation of *A. arenosa* *ATHILA* TEs is associated with non-CG hypomethylation in F1 hybrids.

Previous analyses of F1 hybrids from intraspecies crosses between C24 and *Ler Arabidopsis thaliana* accessions have revealed non-additive DNA methylation inheritance at numerous loci and a global decrease in 24-nt sRNA accumulation (Groszmann et al. 2011a, Greaves et al. 2012a). However, in this case, most of the differences between hybrids and parents occurred at genic and gene-proximal regions, in sharp contrast with the situation in F1 epihybrids, in which the majority of methylation alterations occurs at TEs. Additionally, no transgressive DNA methylation patterns were detected in C24xLer hybrids. This suggests that the epigenomic shock may only be triggered when parent plants show marked differences in DNA methylation level, notably at TEs. This is likely not the case between C24 and Ler as previous studies have pointed out that transposon sequences are almost invariably methylated in wild *Arabidopsis* accessions (Vaughn et al. 2007b, Schmitz et al. 2013).

Various stresses or environmental conditions can modulate patterns of methylation and/or activate TE transcription (Tittel-Elmer et al. 2010, Downen et al. 2012a). Further, mutations altering the function of DNA methylation regulators may spontaneously arise in natural populations, generating individuals with globally hypomethylated genomes and activated TEs. Although *Arabidopsis thaliana* is a mainly self-fertilizing species, its effective outcrossing rate varies considerably between different stands and has been reported to reach more than 14% (Bomblies et al. 2010a). Thus, it is plausible that epigenomic shocks, though perhaps less drastic than the one we report here, may occur in the wild and contribute to generate novel heritable epigenetic diversity and associated regulatory patterns. Given the impact of such shocks on TEs, this may also have important implications for genomic TE dynamics.



## Materials and Methods

### Plant material

Isolation of the *met1-3* mutant has been described previously (Saze et al. 2003). The original *met1-3* heterozygote was backcrossed twice in the heterozygous state. First generation *met1-3* mutant lines were selected from self-pollinated progeny of the heterozygote, and second-generation *met1-3* mutants obtained from self-pollination of first generation *met1-3* mutants were used in this study. Plants were grown on soil under long-day conditions (16 hr light/8 hr dark).

### RNA sequencing

Libraries for RNA sequencing were prepared from 2 µg of total RNA from leaf tissues using the TruSeq RNA sample prep kit from Illumina according to the manufacturer's protocol. RNA libraries were sequenced on an Illumina HiSeq2000 instrument with 2 x 101 bp paired-end reads. Samples were randomly combined into pools of 6 per lane. For image analysis and base calling, we used the Illumina OLB software version 1.8. Reads were quality-trimmed and quality-filtered with the SHORE pipeline v0.9.0 (Ossowski et al. 2008). Adapter sequences were trimmed using trimReads (<https://github.com/tanghaibao/trimReads>). We identified pairs having dovetail overlaps with sortPairedReads, using the default settings (<https://github.com/tanghaibao/trimReads>). We used FLASH with default settings to merge the overlapping reads into extended single reads (Magoc et al. 2011). Reads were mapped following two independent strategies. For one, reads were mapped to the TAIR9 version of the *A. thaliana* genome (<http://www.arabidopsis.org>) using Tophat v2.0.8 with Bowtie v2.1.0 (Langmead et al. 2012, Kim et al. 2013). We retained only uniquely mapping reads. The *read-realign-edit-dist*, *micro-exon-search*, *coverage-search* and *no-mixed* (for paired-read mapping) modes were enabled. Paired reads were mapped first; identified junctions were then fed into the single read mapping. In the second approach, reads were mapped using bwa v0.6.0 (Li et al. 2009) against the TAIR10 version of the *A. thaliana* gene and transposon annotation (<http://www.arabidopsis.org>). Only uniquely mapping reads were retained; in the case of paired reads, only concordant pairs were considered.

### Gene expression analysis

We applied two independent strategies. In a first approach, we used Cufflinks version 2.0.2 (Trapnell et al. 2013) for the quantification of gene expression.



We ran a Reference Annotation Based Transcript assembly (RABT) using the TAIR10 gene and transposon annotations. Ribosomal RNA genes, plastidic and mitochondrial sequences were masked from the all analyses. *Fragment-bias-correction*, *multi-read correction*, *total-hits-normalization* and *upper-quartile-normalization* modes were enabled. Maximum intron length was set to 20,000 bp. Transcripts from each sample were merged using Cuffmerge v2.0.2, with RABT enabled. For detection of differential gene expression we ran Cuffdiff v2.0.2 on the merged transcripts (FDR < 0.05). The *fragment-bias-correction*, *multi-read-correction* and *upper-quartile-normalization* modes were enabled.

In the second approach reads or read pairs mapping to genes and transposable elements according to the TAIR10 annotation were counted using a custom perl script. Gene expression was quantified using the DESeq package (Anders and Huber, Genome Biol. 2010). Differentially expressed genes were identified by a combination of per-gene variance ( $P < 0.05$  after Benjamini-Hochberg correction) and common variance ( $> 2x$  fold change).

### **Whole-genome bisulfite sequencing**

Five  $\mu$ g genomic DNA from leaf tissues was used to generate BS-seq directional libraries using the MethylC-Seq protocol (Lister et al. 2008). The bisulfite treatment was performed using ZYMO EZ DNA Methylation Gold kit. Libraries were sequenced by the BGI on a HiSeq 2000, producing 90-mer paired reads. Identical pairs were collapsed and unpaired reads were removed. Libraries were mapped to the TAIR10 genome using the BS\_seeker2 pipeline (Guo et al. 2013), with the Bowtie2 aligner (Langmead et al. 2012) in local alignment mode, allowing 4 mismatches and a maximum insert size of 450 bp. BS\_seeker2 output was filtered to conserve only cytosines with a minimum coverage of 6 reads.

### **Identification of DMRs**

DMRs were determined as previously described (Stroud et al. 2013), using the following threshold: 0.3, 0.2, 0.16 for CG, CHG, CHH in the *met1/Col*, *ibm1/WT* and *rdd/WT* comparisons; 0.15, 0.1, 0.08 for CG, CHG, CHH in the epihybrids / rF1 comparisons.

### **Metagene plots of DNA Methylation**

We used a custom perl script to compute the average methylation pattern of a set of loci. For each locus, the script creates 50 bins: 20 bins of 100 bp each for the 2kb upstream and 2kb downstream of the locus, and ten bins of equal length corresponding to the locus accession. The methylation level of each bin of a given locus was computed. The average methylation level for each bin in all loci of the set was plotted.





### **Small RNA sequencing**

Libraries for small RNA sequencing were prepared from 3 µg of the same total RNA samples used for RNA sequencing. Sequencing libraries were produced and sequenced by Fasteris on an Illumina HiSeq 2500. Reads were mapped to the Arabidopsis TAIR10 genome using TopHat allowing no mismatches. The number of 21-nt and 24-nt uniquely mapped reads corresponding to each non-overlapping 100-bp genomic windows was calculated and expressed as normalized RP10M values (reads per 10 millions of total 18-26-nt reads of the whole library).

### **Region of differential small RNA accumulation**

To determine regions of differential small RNA (21- or 24-nt) accumulation between 2 samples, we applied a Fisher's Exact Test to the normalized RP10M values of each 100-bp genome windows containing at least 4 mapped reads in one of the 2 samples. Windows with a Benjamini-Hochberg corrected p-value lower than 0.01 and an absolute fold-change greater than 4 (Col versus *met1*) or 2 (F1 epihybrids versus rF1) were deemed as region of differential small RNA accumulation.

### **RT-PCR assays**

Semi-quantitative RT-PCR reactions were performed from total leaf RNA using the OneStep RT-PCR kit (Qiagen) as described previously (Rigal et al. 2012). A derived-CAPS (dCAPS) assay was used to distinguish allele-specific IBM1-L transcript accumulation in Ler-0 x *met1*-3 hybrids (Rigal et al. 2012). Quantitative RT-PCR reactions were performed from total leaf RNA using SensiFAST™ SYBR No-ROX One-Step Kit (Bioline) in an Eco Real-time PCR system (illumina).

### **Locus-specific DNA methylation analyses**

Bisulfite conversion and sequencing of cloned PCR products were performed as described previously (Rigal et al. 2012). To study *IBM1* intron methylation, before PCR amplification, genomic DNA was first digested overnight with 5 units of *DraI* followed by a 3-hour *AluI* digestion with 20 units of enzyme and final heat inactivation. McrBC measurement of DNA methylation was performed starting from 250 ng of genomic DNA as described in (Teixeira et al. 2009). PCR amplification of digested templates were performed using the GoTaq polymerase (Promega), or by qPCR using the SensiFAST™ SYBR No-Rox Kit (Bioline).

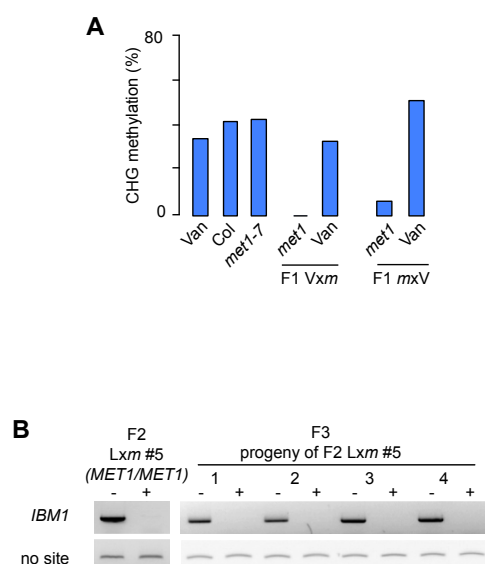


**Acknowledgements**

We thank Jerzy Paszkowski for sharing unpublished data. This work was supported by the CNRS (O.M.), the European Community's Seventh Framework Programme through a Starting Independent Research Grant from the European Research Council (I2ST, 260742; O.M.), a Young Researcher grant from the Auvergne Regional Council (O.M.), the Max Planck Society (D.W.) and the Deutsche Forschungsgemeinschaft (SFB1101; D.W.) O.M. is an EMBO Young Investigator.

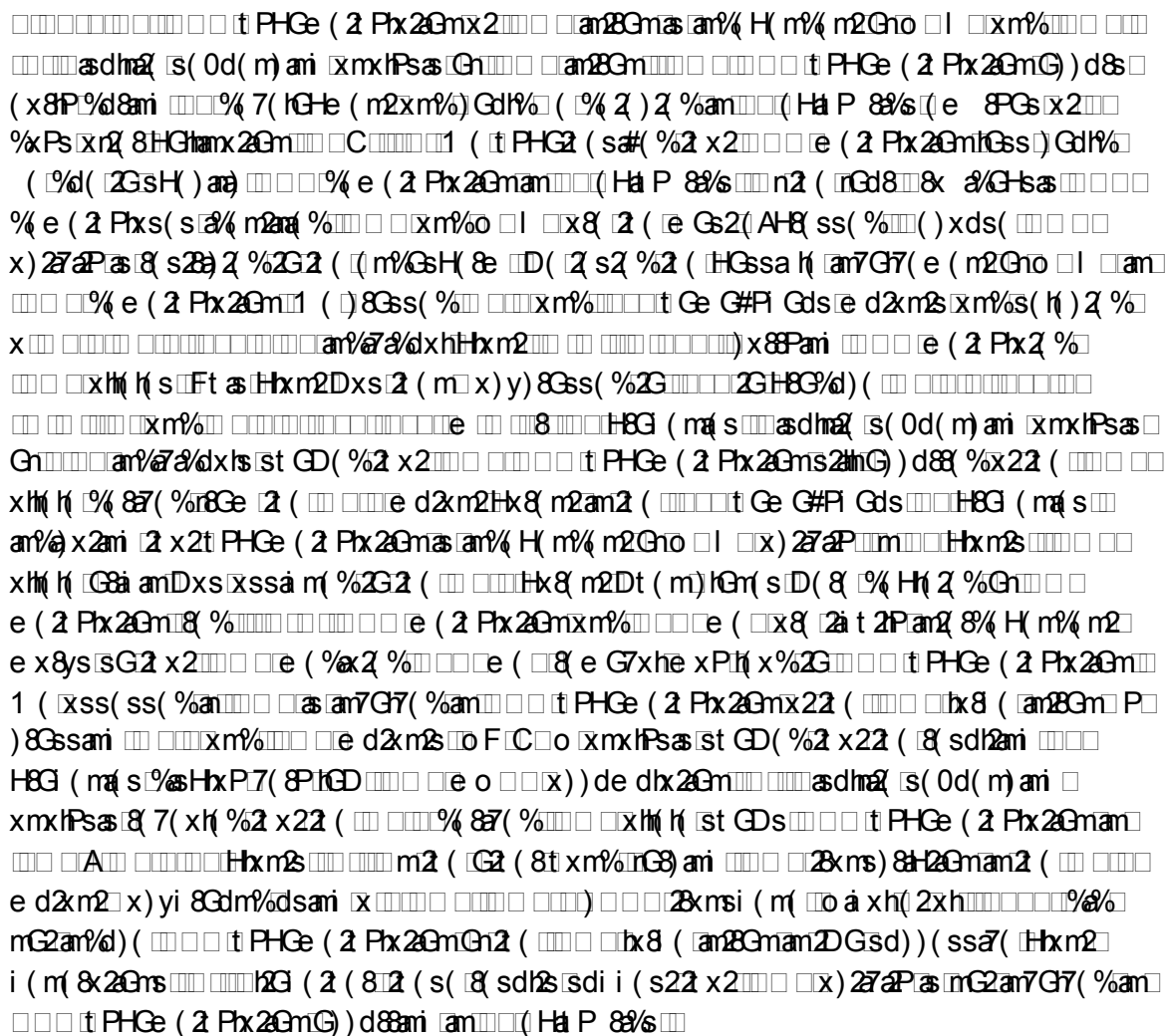


# Supplemental Data



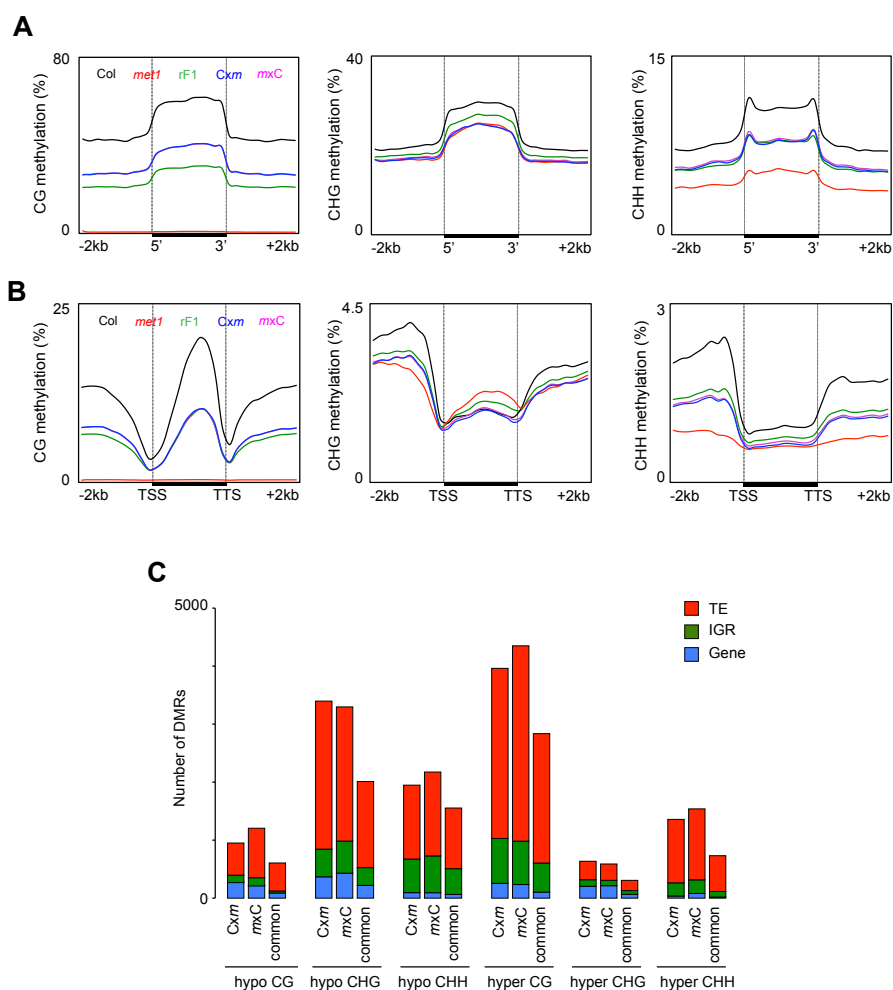
**Figure S1.** CHG hypomethylation at *IBM1* in F1 plants and in self-pollinated progenies. (A) CHG methylation level at the large *IBM1* intron in the indicated genotypes. (B) DNA methylation analysis at the large *IBM1* intron using *A**lu*I digestion (inhibited by CHG methylation) followed by PCR. For each genotype, PCR amplification was performed on digested (+) and undigested DNA (-). A region deprived of *A**lu*I site was amplified as a control of DNA loading.









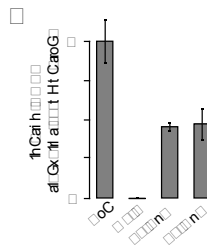
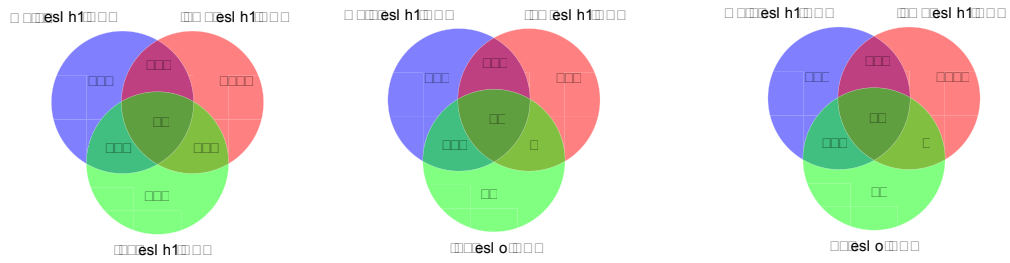


**Figure S3.** DNA methylation changes in epihybrids.

(A) Average distribution of CG, CHG and CHH methylation along all TEs (over 1kb), and along all genes (B) in indicated genotypes. (C) Number of F1 DMRs of the indicated type defined in each epihybrid relative to rF1 and common to both epihybrids. F1 DMRs overlapping with TEs, intergenic regions (IGR) and genes are represented in red, green and blue, respectively.



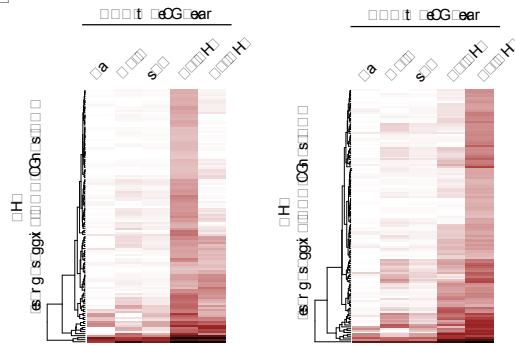
□



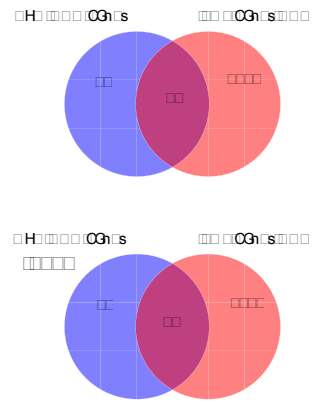
Geh1ra G h oyHha rG t ch esi h1HæsCaoG  
 hGG r p1Hx oyphGx oi h1Cl rGp v ræ rG r an x F G Gx r oy  
 al G x il aoGrG l hGa G hl res r C Gx 1b1 k hl 1hxGaæh xa G 1 h1b1oy  
 æhHh GybH æ1hh rG hl hG hGa roOpr Chl C æhx



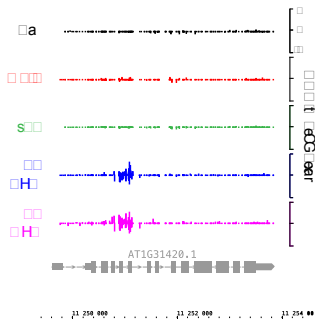
□



□

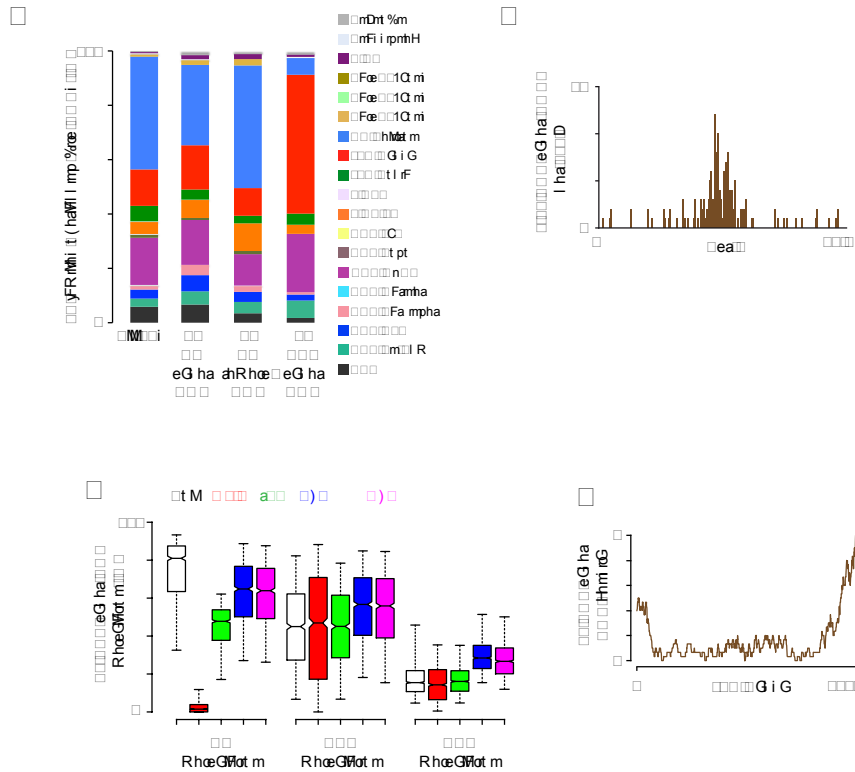


□



asrgs ggt CGst eG ear x nxCG sg  
et na t eG ear i gvx asrgs ggt CGst eG ear g  
x x H e r H s CG sg at savgsixv a Ht n  
r gCav x anx CGst eG ear x nxCG sg r r x st ga r g  
ai s nnx vx x e g

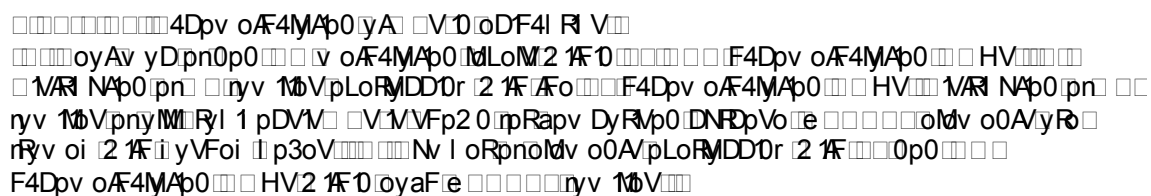




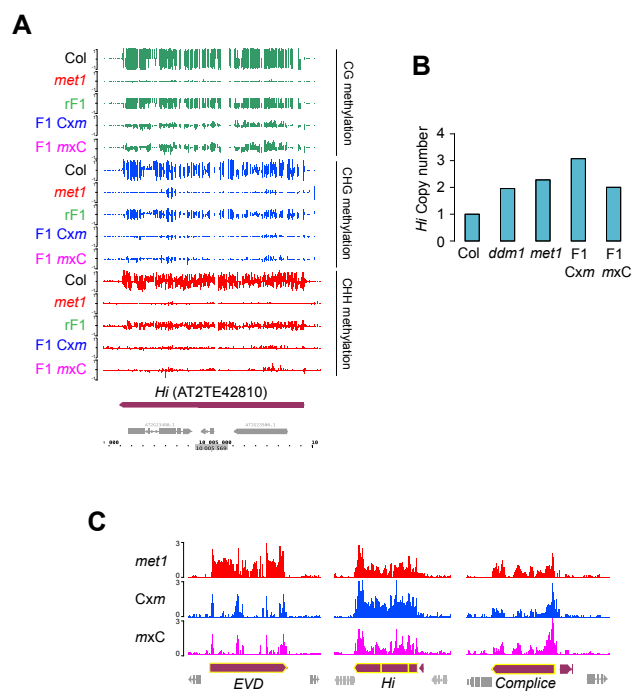
FrH eG haRhaeGha m m h l reG aH  
 ri æ not mty yFRmhi t (haMl l mp %ce æh mHCFhH i ri æ not mty  
 yFRmhi t yFMhaF rH l i ri i et %mxt aQ RI Fai t m l na t i h ed Rt i t RFM rh %t y  
 æh ChRnMæ(hmR hat y eG haRhaeGha m m i m D oMi FMmp  
 Cæd Rt i t Rh Mot y FrH RhaeGha m m (hM m  
 eG haRhaeGha hH i OFMhi %hah) æFChHyt R t m h l MFCh haphmt æ h  
 ri æ not mty eG t RhaeGha m m Hmi æGt (ha G i GhMRhro





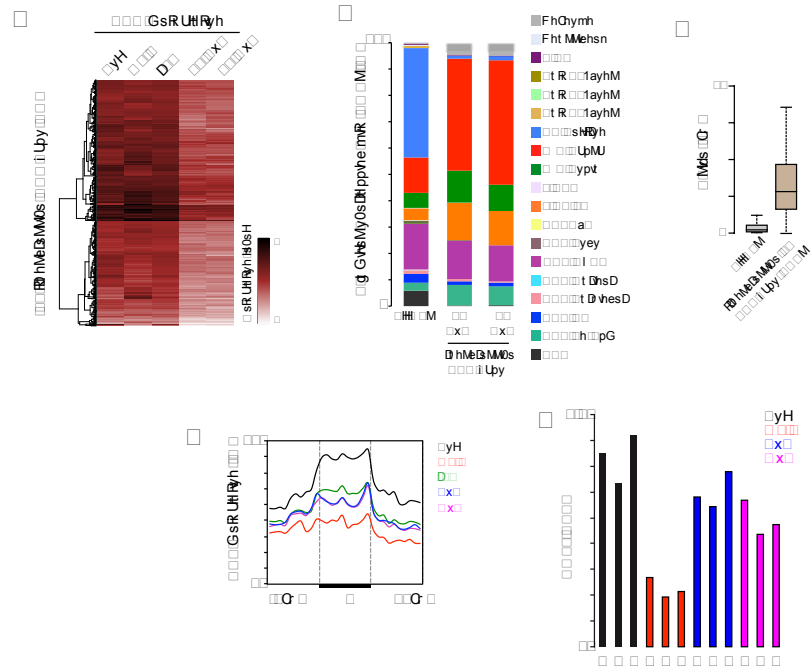






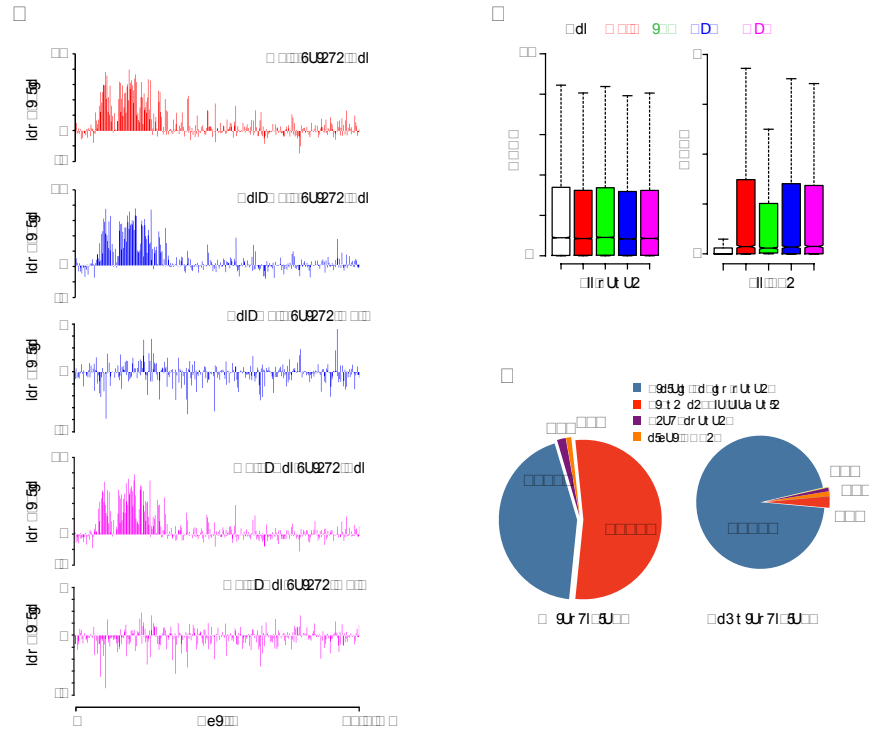
**Figure S8.** *Hi* hypomethylation and copy number of *Hi*, *EVD* and *Complice*  
 (A) Genome-browser view of DNA methylation at the *Hi* element. (B) *Hi* copy number in the indicated genotypes as determined by quantitative PCR. (C) Ratio of methyl-C DNA sequencing coverage in the indicated genotypes over sequencing coverage in the Col parent.





D hMeBMMOs i UpyGsR UTRh v hspv U r DnM  
 st RGt pyg GsR UTRh hOsHmR h D hMeBMMOs i UpyGsR UTRh M  
 MR I Ryhyg g GsMyOsDhpphe mR D hMeBMMOs i UpyGsR UTRh  
 M MR I Ryhyg g GsMygt H D r nypMM M MM ymh g DayGpt D yh pl D yM  
 yx pl Ryg hsheR ayGpt Dhe t H D r nypMM Mt hn MyOsDhpphe mR  
 D hMeBMMOs i UpyGsR UTRh M OsDes GsR UTRh t hhe M  
 yOsDhpphe mR D hMeBMMOs i UpyGsR UTRh M sOsHyg G  
 t aal GI H Ryh h vnat R neshyRpsMont R gY G Rs Mo D phat R Mt B M ymh

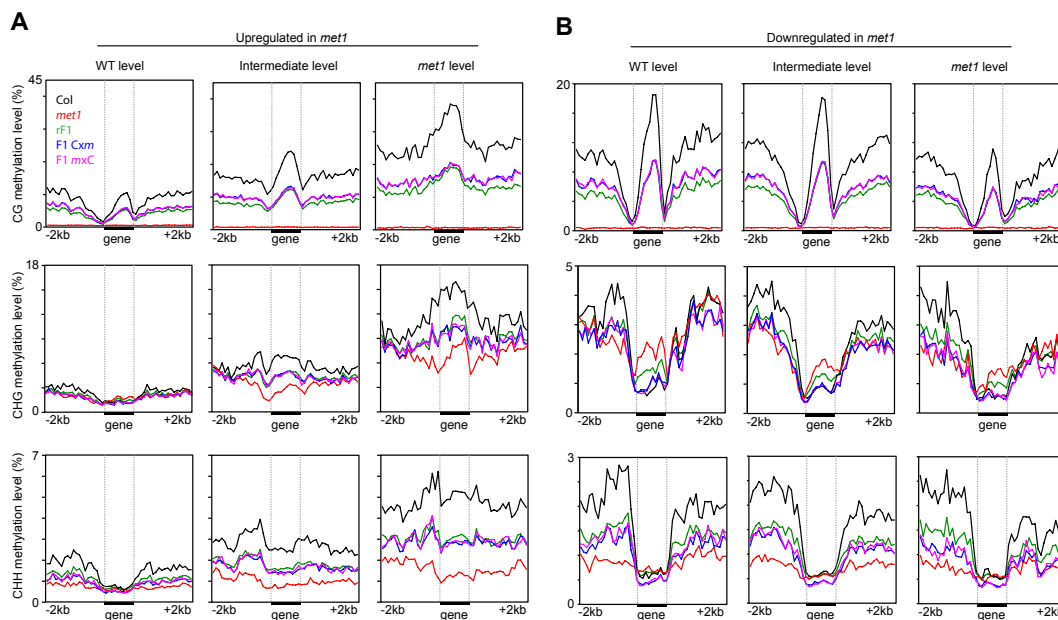




2U1 t lo22 dpr l 5 t 2 9 5 t e t r U2 g U 9 2 l t 2  
 e9da d2da 6g3 dp5U 6U9 r Uldr dp g g 5U 9 5 g u 5U2 ldt r  
 e9da d2da U dD ld52 dp 7a 7l 5 t U6U2 dp l r Ut U2 t l l 2 g  
 g g 5U 2 a lU2 d5 l t 7a U9 t 25g 75 t dpa 2 9r 7l 5U ld gg 5eU a 75 t 5

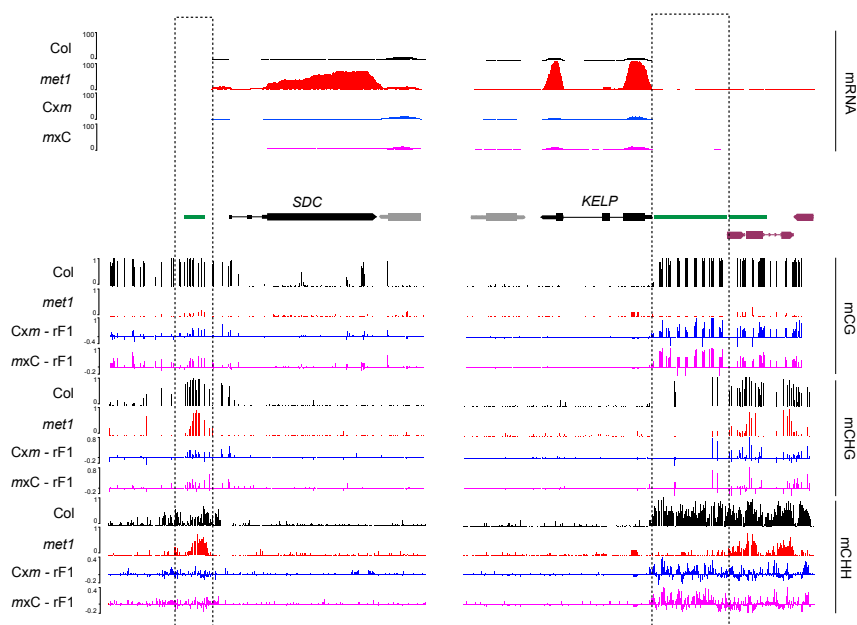






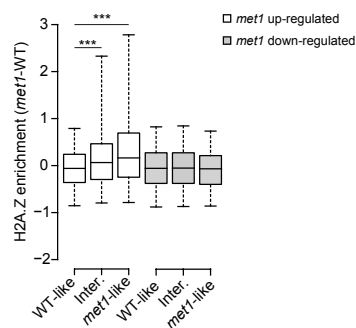
**Figure S11.** Average DNA methylation profiles of *met1*-misregulated genes grouped according to their transcriptional level in F1 epihybrids. (A) Average distribution of CG, CHG and CHH methylation along all genes up-regulated in *met1*. (B) Average distribution of CG, CHG and CHH methylation along all genes down-regulated in *met1*. WT level, Intermediate level and *met1* level refers to the transcription level in F1 epihybrids.





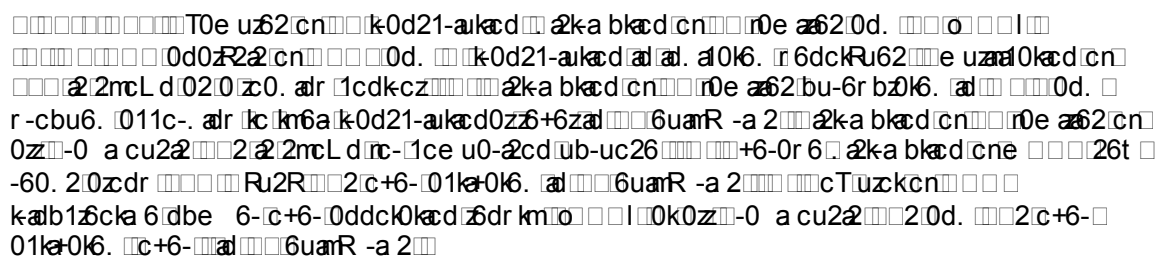
**Figure S12.** Genome-browser view of DNA methylation and mRNA levels at the *SDC* and *KELP* loci. Genome-browser view of DNA methylation and mRNA levels at the *SDC* and *KELP* loci. The F1 remethylated repeats in the promoters of *SDC* and *KELP* are framed with a dashed box. *SDC* and *KELP* genes are in black, other genes are in grey, TEs are in purple, repeats are in green.





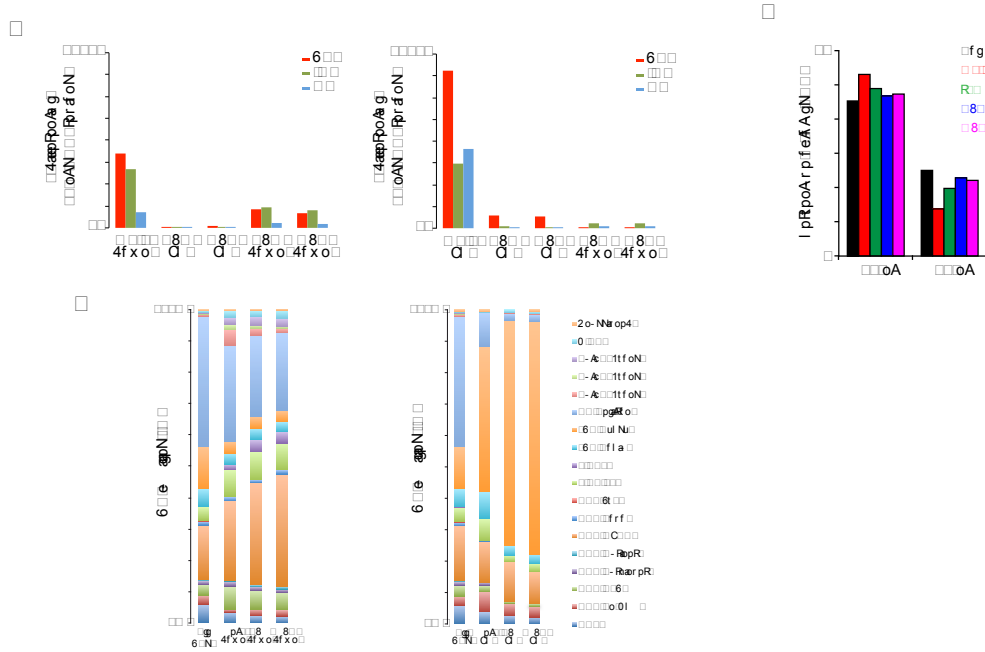
**Figure S13.** Genes activated in *met1* and transcribed in F1 are enriched in H2A.Z in *met1* Box plot of H2A.Z enrichment over genes up- and down-regulated in *met1* grouped according to their transcriptional level in F1 epihybrids. H2A.Z enrichment values in the WT and in *met1-6* were extracted from public data (Zilberman et al, 2008). (Inter., Intermediate). \*\*\*  $P < 10^{-4}$  by two-tailed Wilcoxon rank-sum test.





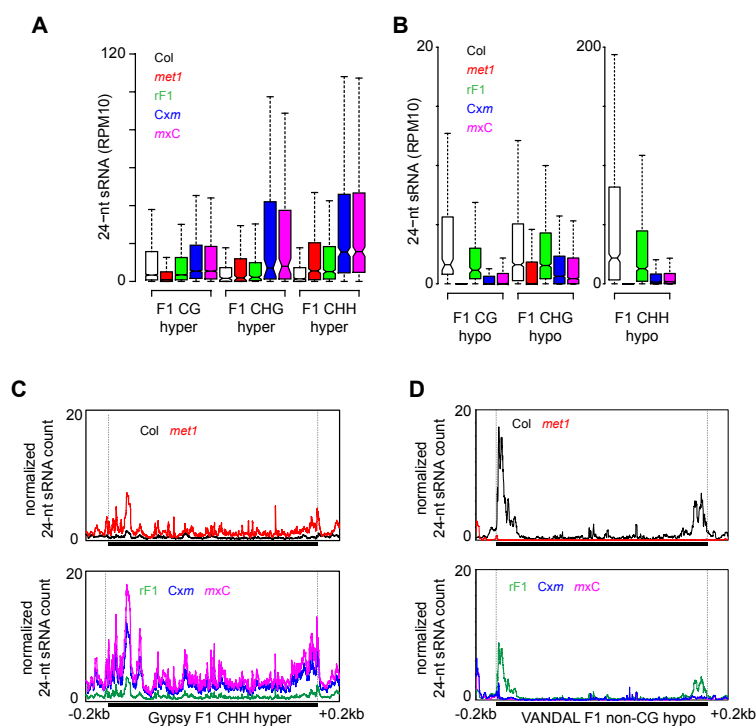






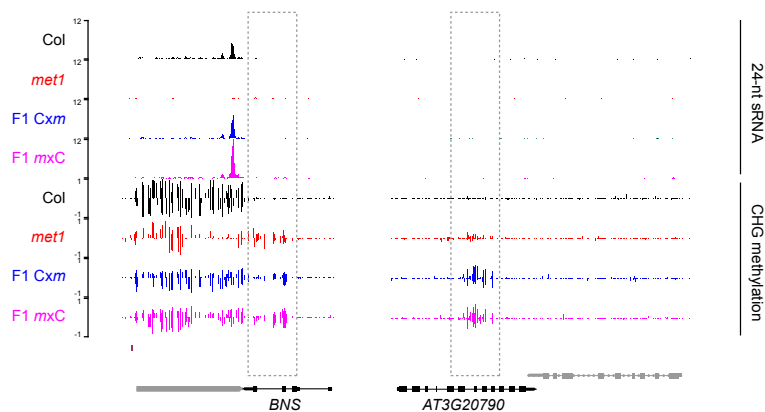
Nps o g Nf e o AN t t C Cg Af o  
 C npRf ePr d oNf e4apPoA g oAN t t C Cg Af o x ac 4pt Rp- Np4 N gFpg  
 4f x o g eAf Rat Rp- Np4 N gFpg Cl Pr cA Af Np Pr d oN  
 Ap gFpgf e oAN a 8 f R 8 pl acun R Nx- N Apotf I- Rp4 x ac R o4 Pr d oN  
 Ncf x ar at Rp- Np4 8 Cl o4 8 Cl f R4pt Rp- Np4 8 4f x o o4 8 4f x o oA  
 N gFpgNx pRp N f Rp4 6cpf FpRg I f eApNp Pr d oNx ac 6 aAR poa Pr d o f R  
 r pop a N Ncf x o pRp oAr p f e oA o4 oAN NRp g Af A Ap oC npRf e oA  
 N Nps Qot p4 f Rp- t c r p of Al p a R a C of e6 e ap Nf FpRg I l ar x ac  
 a4a- Ap4 Pr d oNf e4apPoA g oAN t t C Cg Af o a R a C of e6 e ap Nf e g  
 R na f l N6 N a Ncf x o f Rf I- R a f o l CR f Np





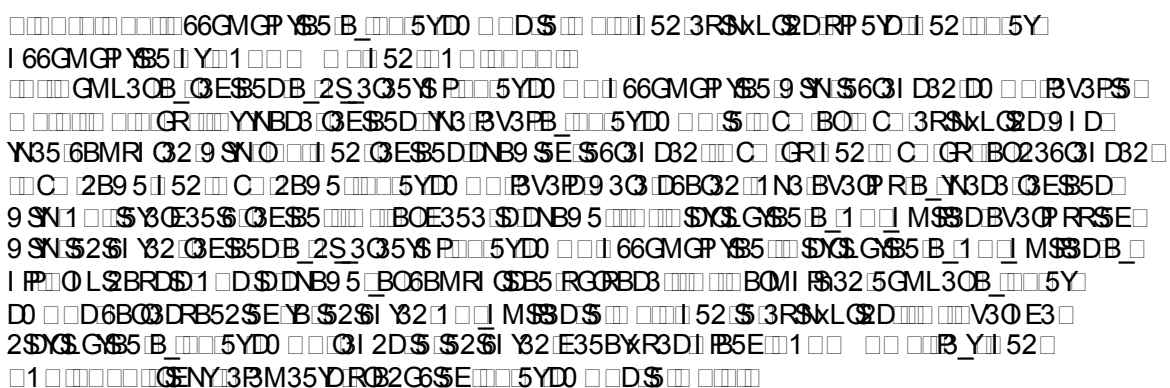
**Figure S16.** Correlation between 24-nt sRNA accumulation and DNA methylation. (A) Box plot of 24-nt sRNA accumulation at F1 hypermethylation DMRs. (B) Box plot of 24-nt sRNA accumulation at F1 hypomethylation DMRs. (C) Average distribution of 24-nt sRNA reads in indicated genotypes along LTR/gypsy elements overlapping with F1 CHH hypermethylation DMRs. (D) Average distribution of 24-nt sRNA reads in indicated genotypes along VANDAL elements overlapping with F1 non-CG hypermethylation DMRs.





**Figure S17.** *BNS* and *AT3G20790* CHG hypermethylation in epihybrids is not associated with accumulation of 24- or 21-nt sRNAs. Genome-browser view of 24-nt sRNA and CHG methylation levels in indicated genotypes at *BNS* and *AT3G20790* loci (black). Other annotations are indicated in grey.









## References

- Aufsatz W, Mette MF, Matzke AJ, Matzke M (2004) The role of MET1 in RNA-directed de novo and maintenance methylation of CG dinucleotides. *Plant Mol Biol* **54**: 793-804. doi:10.1007/s11103-004-0179-1
- Becker C, Hagmann J, Muller J, Koenig D, Stegle O, Borgwardt K, Weigel D (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. *Nature* **480**: 245-249. doi:10.1038/nature10555
- Bomblies K, Yant L, Laitinen RA, Kim ST, Hollister JD, Warthmann N, Fitz J, Weigel D (2010) Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of Arabidopsis thaliana. *PLoS Genet* **6**: e1000890. doi:10.1371/journal.pgen.1000890
- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**: 215-219. doi:10.1038/nature06745
- Coleman-Derr D, Zilberman D (2012) Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. *PLoS Genet* **8**: e1002988. doi:10.1371/journal.pgen.1002988
- Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC, Martienssen RA (2014) miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. *Nature* **508**: 411-415. doi:10.1038/nature13069
- Cui H, Fedoroff NV (2002) Inducible DNA demethylation mediated by the maize Suppressor-mutator transposon-encoded TnpA protein. *Plant Cell* **14**: 2883-2899.
- Deleris A, Stroud H, Bernatavichute Y, Johnson E, Klein G, Schubert D, Jacobsen SE (2012) Loss of the DNA methyltransferase MET1 Induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in Arabidopsis thaliana. *PLoS Genet* **8**: e1003062. doi:10.1371/journal.pgen.1003062
- Dowen RH, Pelizzola M, Schmitz RJ, Lister R, Dowen JM, Nery JR, Dixon JE, Ecker JR (2012) Widespread dynamic DNA methylation in response to biotic stress. *Proc Natl Acad Sci U S A* **109**: E2183-2191. doi:10.1073/pnas.1209329109
- Du J, Johnson LM, Groth M, Feng S, Hale CJ, Li S, Vashisht AA, Gallego-Bartolome J, Wohlschlegel JA, Patel DJ, Jacobsen SE (2014) Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Mol Cell* **55**: 495-504. doi:10.1016/j.molcel.2014.06.009



- Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E, Vashisht AA, Terragni J, Chin HG, Tu A, Hetzel J, Wohlschlegel JA, Pradhan S, Patel DJ, Jacobsen SE (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* **151**: 167-180. doi:10.1016/j.cell.2012.07.034
- Fu Y, Kawabe A, Etcheverry M, Ito T, Toyoda A, Fujiyama A, Colot V, Tarutani Y, Kakutani T (2013) Mobilization of a plant transposon by expression of the transposon-encoded anti-silencing factor. *EMBO J* **32**: 2407-2417. doi:10.1038/emboj.2013.169
- Greaves IK, Groszmann M, Ying H, Taylor JM, Peacock WJ, Dennis ES (2012) Trans chromosomal methylation in Arabidopsis hybrids. *Proc Natl Acad Sci U S A* **109**: 3570-3575. doi:10.1073/pnas.1201043109
- Groszmann M, Greaves IK, Albertyn ZI, Scofield GN, Peacock WJ, Dennis ES (2011) Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. *Proc Natl Acad Sci U S A* **108**: 2617-2622. doi:10.1073/pnas.1019217108
- Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, Chen PY, Pellegrini M (2013) BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* **14**: 774. doi:10.1186/1471-2164-14-774
- Hagmann J, Becker C, Muller J, Stegle O, Meyer RC, Wang G, Schneeberger K, Fitz J, Altmann T, Bergelson J, Borgwardt K, Weigel D (2015) Century-scale methylome stability in a recently diverged Arabidopsis thaliana lineage. *PLoS Genet* **11**: e1004920. doi:10.1371/journal.pgen.1004920
- Henderson IR, Jacobsen SE (2008) Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Genes Dev* **22**: 1597-1606. doi:10.1101/gad.1667808
- Inagaki S, Miura-Kamio A, Nakamura Y, Lu F, Cui X, Cao X, Kimura H, Saze H, Kakutani T (2010) Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome. *EMBO J* **29**: 3496-3506. doi:10.1038/emboj.2010.227
- Jacobsen SE, Meyerowitz EM (1997) Hypermethylated SUPERMAN epigenetic alleles in arabidopsis. *Science* **277**: 1100-1103.
- Jacobsen SE, Sakai H, Finnegan EJ, Cao X, Meyerowitz EM (2000) Ectopic hypermethylation of flower-specific genes in Arabidopsis. *Curr Biol* **10**: 179-186.
- Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Bulski A, Albuissou J, Heredia F, Audigier P, Bouchez D, Dillmann C, Guerche P, Hospital F, Colot V (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet* **5**: e1000530. doi:10.1371/journal.pgen.1000530



- Josefsson C, Dilkes B, Comai L (2006) Parent-dependent loss of gene silencing during interspecies hybridization. *Curr Biol* **16**: 1322-1328. doi:10.1016/j.cub.2006.05.045
- Kakutani T, Munakata K, Richards EJ, Hirochika H (1999) Meiotically and mitotically stable inheritance of DNA hypomethylation induced by ddm1 mutation of *Arabidopsis thaliana*. *Genetics* **151**: 831-838.
- Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, Riddle NC, Verbsky ML, Richards EJ (2003) *Arabidopsis* MET1 cytosine methyltransferase mutants. *Genetics* **163**: 1109-1122.
- Kato M, Miura A, Bender J, Jacobsen SE, Kakutani T (2003) Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*. *Curr Biol* **13**: 421-426.
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: R36. doi:10.1186/gb-2013-14-4-r36
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359. doi:10.1038/nmeth.1923
- Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11**: 204-220. doi:10.1038/nrg2719
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760. doi:10.1093/bioinformatics/btp324
- Lippman Z, May B, Yordan C, Singer T, Martienssen R (2003) Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol* **1**: E67. doi:10.1371/journal.pbio.0000067
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**: 523-536. doi:10.1016/j.cell.2008.03.029
- Magoc T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**: 2957-2963. doi:10.1093/bioinformatics/btr507
- Mathieu O, Probst AV, Paszkowski J (2005) Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in *Arabidopsis*. *EMBO J* **24**: 2783-2791. doi:10.1038/sj.emboj.7600743



- Mathieu O, Reinders J, Caikovski M, Smathajitt C, Paszkowski J (2007) Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. *Cell* **130**: 851-862. doi:10.1016/j.cell.2007.07.007
- Matsushita Y, Deguchi M, Youda M, Nishiguchi M, Nyunoya H (2001) The tomato mosaic tobamovirus movement protein interacts with a putative transcriptional coactivator KELP. *Mol Cells* **12**: 57-66.
- Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, Cao J, Weigel D, Paszkowski J, Mathieu O (2009) Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* **461**: 427-430. doi:10.1038/nature08328
- Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T (2001) Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. *Nature* **411**: 212-214. doi:10.1038/35075612
- Nuthikattu S, McCue AD, Panda K, Fultz D, DeFraia C, Thomas EN, Slotkin RK (2013) The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiol* **162**: 116-131. doi:10.1104/pp.113.216481
- Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D (2008) Sequencing of natural strains of Arabidopsis thaliana with short reads. *Genome Res* **18**: 2024-2033. doi:10.1101/gr.080200.108
- Reinders J, Wulff BB, Mirouze M, Mari-Ordonez A, Dapp M, Rozhon W, Bucher E, Theiler G, Paszkowski J (2009) Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. *Genes Dev* **23**: 939-950. doi:10.1101/gad.524609
- Rigal M, Kevei Z, Pelissier T, Mathieu O (2012) DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns. *EMBO J* **31**: 2981-2993. doi:10.1038/emboj.2012.141
- Rigal M, Mathieu O (2011) A "mille-feuille" of silencing: epigenetic control of transposable elements. *Biochim Biophys Acta* **1809**: 452-458. doi:10.1016/j.bbagr.2011.04.001
- Sasaki T, Kobayashi A, Saze H, Kakutani T (2012) RNAi-independent de novo DNA methylation revealed in Arabidopsis mutants of chromatin remodeling gene DDM1. *Plant J* **70**: 750-758. doi:10.1111/j.1365-3113.2012.04911.x
- Saze H, Kakutani T (2007) Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1. *EMBO J* **26**: 3641-3652. doi:10.1038/sj.emboj.7601788





Saze H, Kitayama J, Takashima K, Miura S, Harukawa Y, Ito T, Kakutani T (2013) Mechanism for full-length RNA processing of Arabidopsis genes containing intragenic heterochromatin. *Nat Commun* **4**: 2301. doi:10.1038/ncomms3301

Saze H, Mittelsten Scheid O, Paszkowski J (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat Genet* **34**: 65-69. doi:10.1038/ng1138

Saze H, Shiraishi A, Miura A, Kakutani T (2008) Control of genic DNA methylation by a jmjC domain-containing protein in Arabidopsis thaliana. *Science* **319**: 462-465. doi:10.1126/science.1150987

Schlappi M, Raina R, Fedoroff N (1994) Epigenetic regulation of the maize Spm transposable element: novel activation of a methylated promoter by TnpA. *Cell* **77**: 427-437.

Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, Schork NJ, Ecker JR (2011) Transgenerational epigenetic instability is a source of novel methylation variants. *Science* **334**: 369-373. doi:10.1126/science.1212959

Schmitz RJ, Schultz MD, Urich MA, Nery JR, Pelizzola M, Libiger O, Alix A, McCosh RB, Chen H, Schork NJ, Ecker JR (2013) Patterns of population epigenomic diversity. *Nature* **495**: 193-198. doi:10.1038/nature11968

Singer T, Yordan C, Martienssen RA (2001) Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). *Genes Dev* **15**: 591-602. doi:10.1101/gad.193701

Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijo JA, Martienssen RA (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**: 461-472. doi:10.1016/j.cell.2008.12.038

Soppe WJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, Peeters AJ (2000) The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol Cell* **6**: 791-802.

Steimer A, Amedeo P, Afsar K, Fransz P, Mittelsten Scheid O, Paszkowski J (2000) Endogenous targets of transcriptional gene silencing in Arabidopsis. *Plant Cell* **12**: 1165-1178.

Stokes TL, Kunkel BN, Richards EJ (2002) Epigenetic variation in Arabidopsis disease resistance. *Genes Dev* **16**: 171-182. doi:10.1101/gad.952102



Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol* **21**: 64-72. doi:10.1038/nsmb.2735

Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* **152**: 352-364. doi:10.1016/j.cell.2012.10.054

Stroud H, Hale CJ, Feng S, Caro E, Jacob Y, Michaels SD, Jacobsen SE (2012) DNA methyltransferases are required to induce heterochromatic re-replication in Arabidopsis. *PLoS Genet* **8**: e1002808. doi:10.1371/journal.pgen.1002808

Tanurdzic M, Vaughn MW, Jiang H, Lee TJ, Slotkin RK, Sosinski B, Thompson WF, Doerge RW, Martienssen RA (2008) Epigenomic consequences of immortalized plant cell suspension culture. *PLoS Biol* **6**: 2880-2895. doi:10.1371/journal.pbio.0060302

Teixeira FK, Heredia F, Sarazin A, Roudier F, Boccara M, Ciaudo C, Cruaud C, Poulain J, Berdasco M, Fraga MF, Voinnet O, Wincker P, Esteller M, Colot V (2009) A role for RNAi in the selective correction of DNA methylation defects. *Science* **323**: 1600-1604. doi:10.1126/science.1165313

Tittel-Elmer M, Bucher E, Broger L, Mathieu O, Paszkowski J, Vaillant I (2010) Stress-induced activation of heterochromatic transcription. *PLoS Genet* **6**: e1001175. doi:10.1371/journal.pgen.1001175

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **31**: 46-53. doi:10.1038/nbt.2450

Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, Miura A, Kakutani T (2009) Bursts of retrotransposition reproduced in Arabidopsis. *Nature* **461**: 423-426. doi:10.1038/nature08351

Vaughn MW, Tanurdzic M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, Colot V, Doerge RW, Martienssen RA (2007) Epigenetic natural variation in Arabidopsis thaliana. *PLoS Biol* **5**: e174. doi:10.1371/journal.pbio.0050174

Vongs A, Kakutani T, Martienssen RA, Richards EJ (1993) Arabidopsis thaliana DNA methylation mutants. *Science* **260**: 1926-1928.

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**: 193-205. doi:10.1016/j.cell.2013.02.033



Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker JR (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* **126**: 1189-1201. doi:10.1016/j.cell.2006.08.003

Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S (2008) Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456**: 125-129. doi:10.1038/nature07324

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* **39**: 61-69. doi:10.1038/ng1929



## IV. Discussion générale

La méthylation CG a un rôle fondamental dans le silencing des éléments transposables mais est également retrouvée au niveau de très nombreux gènes et participerait à maintenir l'expression correcte de ces gènes. Dans le mutant *met1*, en sus de la dérégulation de transposons, la perte de méthylation CG est associée à l'apparition ectopique de méthylation non-CG sur de nombreux loci (Zhang et al. 2006, Mathieu et al. 2007, Cokus et al. 2008, Lister et al. 2008). La réduction de l'expression des déméthylases ADN, dont *ROS1*, ainsi qu'un ciblage incorrect de la voie RdDM contribuent à cette méthylation ectopique (Mathieu et al. 2007). Cependant, ces deux observations n'expliquent pas l'hyperméthylation du corps de nombreux gènes en contexte CHG des cytosines. Parallèlement, dans le mutant *met1*, la marque histone répressive H3K9me2, largement associée à la méthylation CHG, est relocalisée, depuis l'hétérochromatine vers l'euchromatine riche en gènes (Tariq et al. 2003).

Une piste quant à l'origine de cette énigmatique hyperméthylation CHG génique nous a été fournie par l'équipe du Pr Kakutani ayant identifié et caractérisé le mutant *ibm1* (Miura et al. 2009, Inagaki et al. 2010a). Dans ce mutant, la perte de fonction de la déméthylase histone H3K9me2 IBM1 engendre, comme dans *met1*, une apparition d'H3K9me2, liée à de la méthylation CHG, dans le corps de nombreux gènes. Ces données nous ont incité à analyser la transcription du gène *IBM1* dans le mutant *met1*. Ce gène présente un intron de longueur atypique (>2kb) qui contient une région fortement méthylée, non seulement sur les positions CG comme la plupart des gènes, mais également en contexte CHG. *IBM1* produit deux transcrits : une forme longue *IBM1-L*, comportant le domaine JmJc impliqué dans l'activité histone déméthylase, ainsi qu'une forme courte, s'arrêtant au début du grand intron. Nous avons démontré que la méthylation CG et CHG dans le grand intron est nécessaire à la transcription correcte de la forme longue. De plus, l'introduction de la construction *pIBM1::IBM1-L* dans le mutant *met1* a permis de restaurer les profils de méthylation sauvages au niveau des gènes cibles d'IBM1 étudiés. Certains de ces gènes, dont la transcription était altérée dans *met1*, retrouvent également un taux d'expression sauvage dans les plantes transformées, parmi eux, le gène codant la déméthylase ADN *ROS1*.

Ainsi, nos résultats dévoilent une nouvelle fonction de la méthylation CG : elle maintient des profils corrects d'H3K9me2 et de méthylation CHG dans les gènes en régulant la transcription du gène *IBM1*. Elle protège donc les gènes de l'apparition ectopique de ces marques répressives et préserve de ce fait leur expression.





Comme dans le cas de la transcription du gène *ROSI*, la méthylation CG contrôle paradoxalement par ce biais une voie permettant d'enlever la méthylation ADN. La méthylation CG agit donc en « chef d'orchestre », assurant le maintien d'une harmonie entre expression des gènes et silencing de l'hétérochromatine.

Trois études subséquentes ont permis de progresser dans la compréhension de la régulation du gène *IBM1*. Ainsi, elles ont démontré que la région fortement méthylée dans le grand intron du gène *IBM1* comprend un domaine hétérochromatique. Les protéines IBM2/AS11/SOG1 et EDM2, en se liant à ce domaine et aux transcrits *IBM1*, favoriseraient la polyadénylation distale des transcrits *IBM1* en empêchant l'utilisation du signal de polydénylation contenu dans le grand intron (Saze et al. 2013a, Wang et al. 2013, Coustham et al. 2014, Lei et al. 2014). Ces protéines jouent le même rôle au niveau d'autres gènes comportant, comme *IBM1*, des domaines intragéniques hétérochromatiques, permettant ainsi la production de transcrits pleine longueur à partir de ces gènes.

Durant l'étude du contrôle de la transcription du gène *IBM1*, nous avons réalisé le rétrocroisement de plantes hypométhylées *met1*, *cmt3* et *kyp* avec des plantes sauvages, afin de déterminer si la restauration de la méthylation corrélait à la restauration d'un taux sauvage d'accumulation de transcrits *IBM1-L*. Nous avons noté que dans les plantes F1 *met1*+/+, si, comme attendu, la méthylation CG n'était pas restaurée sur l'allèle *IBM1* provenant du parent mutant, de façon très surprenante, le taux de méthylation CHG de cet allèle était fortement réduit.

Nous avons montré que la perte complète en F1 de méthylation CG et CHG sur l'allèle *IBM1* dérivé du parent *met1* est ensuite héritée de façon stable à travers la méiose sur au moins deux générations, même en présence de la fonction sauvage MET1. Cet allèle déméthylé est peu transcrit, pouvant ainsi engendrer l'apparition de profils aberrants de méthylation ADN au niveau des gènes cibles d'*IBM1*. Pourquoi la méthylation CHG est-elle maintenue dans l'intron *IBM1* lors de la propagation du mutant *met1* par autofécondation mais perdue lors de rétrocroisement ? Ceci reste à déterminer.

Le locus *IBM1* n'est pas le seul affecté par le rétrocroisement des plantes *met1* puisque le séquençage bisulfite a permis de montrer que les profils de méthylation des épihybrides F1 divergeaient fortement du taux attendu, intermédiaire entre les parents. Ceci est particulièrement vrai au niveau des éléments transposables où nous avons identifié des profils inattendus tels qu'une hyperméthylation CG ou CHH, ou encore la déméthylation partielle voire totale de certains éléments.



Le parent *met1* diffère du parent sauvage non seulement par sa forte déficience en méthylation CG, mais également par les profils de méthylation non-CG et H3K9me2. De plus, de nombreux loci sont différenciellement exprimés dans le mutant. On peut donc envisager que la combinaison de ces différences entre les deux parents joue un rôle dans l'apparition des altérations épigénétiques et transcriptionnelles survenant dans les épihybrides F1. Précédemment, des études ont décrit les profils transcriptionnels et/ou épigénétiques de plantes hybrides F1, issues de croisement entre accessions différentes (C24 x Ler notamment) ou de rétrocroisement d'autres mutants hypométhylés. Dans le premier cas, les deux parents possèdent des génomes proches mais portant tout de même un certain bagage de polymorphisme génétique et épigénétiques (Groszmann et al. 2011b, Greaves et al. 2012b, Shen et al. 2012). De plus, ils possèdent des profils d'expression géniques différents. Dans les hybrides F1 C24 x Ler, des altérations de méthylation et de transcription ont été observées. Cependant, à la différence de nos épihybrides, elles affectent très majoritairement des régions géniques produisant des siRNAs et seule la méthylation CHH semble significativement perturbée (Groszmann et al. 2011b, Greaves et al. 2012b, Shen et al. 2012). Les mécanismes de trans-méthylation décrits par Greaves *et al.* en 2012 sont donc probablement distincts de ceux participant à la restructuration des profils de méthylation dans les épihybrides *met1* x WT où la méthylation CHH génique est très peu impactée. Le rétrocroisement des mutants hypométhylés *cmt3* ou *kyp*, induit, quant à lui, la restauration fidèle des taux de méthylation et de transcription sauvage au niveau des transposons analysés (Lippman et al. 2003a). Ces données suggèrent que les nouveaux profils de méthylation ADN survenant dans les épihybrides *met1/+* sont liés à la perte de la méthylation CG dans le parent mutant plutôt qu'aux différences transcriptionnelles.

La vague de déméthylation d'éléments transposables que l'on a pu noter dans les épihybrides *met1/+*, causant une mobilisation de certains d'entre eux, n'est pas sans rappeler des phénomènes précédemment observés lors de la génération d'hybrides inter-espèces, dont certains hybrides de Marsupiaux (O'Neill et al. 1998). En effet, dans les descendants issus du croisement de wallabies bicolores et de wallabies de l'île Eugène par exemple, une hypométhylation a été détectée sur le génome entier, en association avec l'amplification de rétroéléments et des remaniements chromosomiques non présents dans les deux parents. Toutefois, jamais une restructuration aussi importante des profils de méthylation, affectant tant les gènes que les éléments transposables, telle que celle survenant dans les épihybrides *met1/+* n'a été jusqu'alors mise en évidence.



Les données obtenues au cours de cette thèse contribuent à la compréhension de la stabilité des marques épigénétiques. Ainsi, elles démontrent tout d'abord que la méthylation CG peut, s'il demeure un léger taux de méthylation dans le parent mutant, être à nouveau immédiatement rétablie sur certains loci après restauration de la fonction sauvage MET1. Cette reméthylation pourrait impliquer une activité *de novo* de la méthyltransférase ADN puisqu'elle cible préférentiellement les cytosines en contexte symétrique. Deux articles antérieurs suggèrent une telle activité de la protéine MET1. L'un décrit l'implication de MET1 dans la reméthylation partielle et stochastique des sites CG d'un gène (Zubko et al. 2012). Le second démontre le rôle crucial de MET1 dans le maintien de la méthylation CG induite par des petits ARN au niveau d'un transgène (Aufsatz et al. 2004a). Il est probable que les siRNAs soient également responsables de la reméthylation CG des éléments transposables dans les épihybrides F1. Ces éléments conservent un fort taux d'accumulation de siRNAs de 24nt dans *met1*, parallèlement à la persistance d'un faible taux de méthylation CG donc la méthylation CG de ces loci est en partie indépendante de MET1. Certains de ces loci sont également resilencés en F1. Au niveau de ces éléments, la méthylation CG serait donc fortement dépendante des siRNAs permettant ainsi une restauration immédiate du silencing dès lors que la fonction MET1 est présente. Ce mécanisme de reméthylation pourrait donc être similaire à celui observé dans les plantes issues du croisement *ddm1* × WT, mais plus efficace puisque dans ce cas, la reméthylation est progressive, un taux proche du sauvage n'étant atteint qu'en F5 (Teixeira et al. 2009, Groszmann et al. 2011b, Greaves et al. 2012b, Shen et al. 2012).

Parallèlement, nos résultats révèlent que la méthylation CHG, déposée sur de très nombreux gènes, dans le mutant *met1*, tend à persister dans les épihybrides. Conjointement aux épiallèles causés par l'hypométhylation CG, de nouveaux épiallèles hyperméthylés CHG créés dans *met1* peuvent donc être stablement hérités. A l'instar de la variabilité génétique, ces épiallèles représentent également une source de variabilité héritable, dans ce cas, épigénétique, pouvant engendrer une variabilité phénotypique. Ce potentiel pourrait intéresser les sélectionneurs à la recherche de nouvelles variétés produisant un meilleur rendement ou de meilleures qualités gustatives.

Mais pourquoi ces loci hyperméthylés dans *met1* ne perdent-ils pas cette méthylation en F1 ? Au moins deux hypothèses peuvent être envisagées : (i) le mécanisme de déméthylation CHG ne serait plus fonctionnel ou serait moins efficace dans les épihybrides, ou (ii) la voie de méthylation CHG serait à l'inverse plus efficace.



L'étude du locus *BNS*, hyperméthylé CHG dans le mutant et les plantes F1 nous apporte peut-être quelques éléments de réponse. En effet, nous avons démontré que l'allèle sauvage du gène est hyperméthylé CHG dans les plantes *met1* x Col à un niveau au moins équivalent au mutant *met1*. De plus, dans les plantes F2, dans lesquelles *IBM1* est transcrit au taux sauvage, le locus *BNS* n'est pas déméthylé. *BNS* étant un gène cible d'*IBM1*, ces résultats suggèrent un problème de recrutement de la protéine IBM1 vers ses cibles. Par ailleurs, nos données révèlent que cette méthylation *de novo* apparaissant dans le mutant *met1* peut être transmise en *trans* à l'allèle *BNS* provenant du parent sauvage dans les épihybrides. Les deux épiallèles *BNS* ainsi générés en F1 conservent également leur état hyperméthylé dans les descendants sauvages issus de l'autofécondation des épihybrides, ce qui indique qu'ils ne sont plus ciblés par l'activité IBM1.

Plusieurs études ont démontré l'implication de siRNAs dans les processus de méthylation en *trans*, tels que la reméthylation d'éléments transposables dans les F1 *ddm1*xWT ou les changements de profils de méthylation ADN dans les hybrides d'*A. thaliana* C24 x Ler-0 (Teixeira et al. 2009, Groszmann et al. 2011b, Greaves et al. 2012b, Shen et al. 2012). L'hyperméthylation CHG de *BNS* dans le mutant *met1* et en F1 n'est, elle, pas corrélée à une apparition de siRNAs de 24 nt ou 21 nt. Nous avons donc mis en évidence un mécanisme nouveau, indépendant des voies de silencing par les petits ARN, qui provoquerait l'hyperméthylation de gènes spécifiquement sur les sites CHG.

Aux côtés du gène *BNS*, 110 gènes montrent une hyperméthylation transgressive dans les épihybrides (supérieure aux deux parents). De façon intéressante, nous avons noté que 60% d'entre eux sont des cibles d'*IBM1*. Ces données suggèrent que l'hyperméthylation CHG, qu'elle soit initiée ou non dans le parent *met1*, serait causée par une déficience d'activité ou de recrutement de la protéine IBM1.

De plus, l'étude des profils de méthylation sur les éléments transposables donne à penser que, dans les plantes F1, l'histone déméthylase ciblerait certains transposons transcrits, comme observé dans les plantes *met1::IBM1-L*, et causerait leur forte déméthylation en CHG. Ceci pourrait être lié au fait que ces éléments possèdent une densité importante en sites CHG, qui semble être une caractéristique préférentielle des cibles d'IBM1 (Inagaki et al. 2010a). Ainsi, dans les épihybrides, l'activité d'IBM1 pourrait être réorientée et induire ainsi la création de nombreux épiallèles, tant sur les gènes que sur les éléments transposables.





La baisse d'expression des déméthylases ADN est également responsable de l'hyperméthylation de nombreux loci survenant dans le mutant *met1* (Penterman et al. 2007b). Une étude de la transmission d'épiallèles hyperméthylés CHG lors du rétrocroisement de plantes *rdd* avec des plantes sauvages a démontré que 11/14 gènes perdent cette hyperméthylation dans les plantes F1. Cependant, 3/14 présentent un taux de méthylation intermédiaire entre les parents, indiquant que certains épiallèles hyperméthylés sont hérités. Nos données montrent que la proportion de gènes cibles du trio ROS1 DML2 DML3, et non déméthylés dans les épihybrides F1, est très importante, suggérant que le ciblage de ces gènes par les déméthylases est perturbé dans ces plantes. D'après une étude récente, la protéine ROS4 participerait au guidage de la protéine ROS1 vers ses cibles grâce à sa liaison à des régions denses en sites CG méthylés et pauvres en méthylation H3K4. Dans les épihybrides F1 *met1/+*, les allèles hérités du parent *met1* sont hypométhylés sur les sites CG. L'absence de déméthylation CHG sur de nombreux gènes pourrait donc être liée à un défaut de reconnaissance des loci par la protéine ROS4, altérant ainsi le recrutement de ROS1.

Dans *met1*, à l'opposé de l'hyperméthylation provoquée par la réduction d'expression des déméthylases ADN, une importante hypométhylation non-CG est observée au niveau des séquences internes de nombreux éléments transposables de grande taille, cibles privilégiées de la méthyltransférase ADN CMT2. Dans les plantes F1, les données de séquençage bisulfite et de RNAseq révèlent la persistance d'une hypométhylation non-CG sur un grand nombre de ces éléments, couplée à une forte réactivation transcriptionnelle. Ceci pourrait être lié à la baisse d'expression du gène *CMT2* que nous avons détecté dans le mutant, et qui est héritée dans les épihybrides. Ainsi, la méthylation ADN stabilise les profils de méthylation non-CG via le contrôle de l'expression des gènes *ROS1*, *DME*, et *IBM1* mais aussi de *CMT2*.

Ainsi, nous avons mis en évidence un grand nombre d'éléments transposables déméthylés et fortement transcrits dans les épihybrides F1. Au niveau de certains de ces éléments (VANDAL, CACTA), l'allèle hypométhylé hérité de *met1* est transcrit et peut provoquer, par l'intermédiaire de protéines codées par les éléments, la déméthylation *en trans* de l'allèle WT correspondant et même induire la transposition de l'élément.

Différents types de stress ou de stimuli environnementaux, telles qu'une hausse de température peuvent moduler les profils de méthylation ADN et activer la transcription d'éléments transposables (Tittel-Elmer et al. 2010, Downen et al. 2012b).



De plus, une mutation spontanée altérant la fonction de facteurs contrôlant la méthylation ADN pourrait survenir dans des populations naturelles, et provoquer une hypométhylation du génome et l'activation d'éléments transposables. Bien qu'*Arabidopsis thaliana* soit une espèce principalement autogame, une étude récente a démontré un taux de pollinisation croisée bien supérieur à celui estimé jusqu'alors et pouvant atteindre 15% (Bomblies et al. 2010b). Il est donc possible d'envisager que le phénomène de déméthylation mis en évidence dans nos épihybrides, et impliquant certaines protéines de transposons, puisse survenir dans des populations naturelles lors du croisement de plantes sauvages possédant chacun un lot différent d'éléments transposables exprimés, et engendrer, dans la descendance, des modifications épigénétiques ainsi que la mobilisation de certains éléments.

En conclusion, les données obtenues au cours de cette thèse apportent des connaissances nouvelles concernant les mécanismes de maintenance et de stabilité des profils de méthylation ADN. En révélant une voie de contrôle de l'expression de gènes codant une histone déméthylase et une méthyltransférase ADN, elles confirment le rôle central joué par la méthylation en contexte CG dans la régulation de différentes marques épigénétiques, dont la méthylation non-CG, tant au niveau des gènes que des éléments transposables. De plus, nous avons mis en lumière des mécanismes nouveaux provoquant la forte activation transcriptionnelle, le resiliencing immédiat d'éléments transposables via leur reméthylation, ainsi que la large déméthylation *en trans* de certaines familles de transposons, pouvant conduire à leur mobilisation. L'étude des épihybrides, permet ainsi d'entrevoir des processus ayant pu moduler l'architecture des génomes via la mobilisation d'éléments transposables.



## RÉFÉRENCES BIBLIOGRAPHIQUES

- "Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*." *Nature* (2000). 408(6814): 796-815.
- Agius, F., A. Kapoor and J. K. Zhu "Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation." *Proceedings of the National Academy of Sciences of the United States of America* (2006). 103(31): 11796-11801.
- Alleman, M., L. Sidorenko, K. McGinnis, V. Seshadri, J. E. Dorweiler, J. White, K. Sikkink and V. L. Chandler "An RNA-dependent RNA polymerase is required for paramutation in maize." *Nature* (2006). 442(7100): 295-298.
- Arteaga-Vazquez, M., L. Sidorenko, F. A. Rabanal, R. Shrivistava, K. Nobuta, P. J. Green, B. C. Meyers and V. L. Chandler "RNA-mediated trans-communication can establish paramutation at the b1 locus in maize." *Proceedings of the National Academy of Sciences of the United States of America* (2010). 107(29): 12986-12991.
- Aufsatz, W., M. F. Mette, A. J. Matzke and M. Matzke "The role of MET1 in RNA-directed de novo and maintenance methylation of CG dinucleotides." *Plant molecular biology* (2004a). 54(6): 793-804.
- Aufsatz, W., M. F. Mette, A. J. Matzke and M. Matzke "The role of MET1 in RNA-directed de novo and maintenance methylation of CG dinucleotides." *Plant Mol Biol* (2004b). 54(6): 793-804.
- Autran, D., C. Baroux, M. T. Raissig, T. Lenormand, M. Wittig, S. Grob, A. Steimer, M. Barann, U. C. Klostermeier, O. Leblanc, J. P. Vielle-Calzada, P. Rosenstiel, D. Grimanelli and U. Grossniklaus "Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis." *Cell* (2011). 145(5): 707-719.
- Barbour, J. E., I. T. Liao, J. L. Stonaker, J. P. Lim, C. C. Lee, S. E. Parkinson, J. Kermicle, S. A. Simon, B. C. Meyers, R. Williams-Carrier, A. Barkan and J. B. Hollick "required to maintain repression2 is a novel protein that facilitates locus-specific paramutation in maize." *The Plant cell* (2012). 24(5): 1761-1775.
- Bartee, L., F. Malagnac and J. Bender "Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene." *Genes & development* (2001). 15(14): 1753-1758.
- Beaulieu, J., M. Jean and F. Belzile "The allotetraploid *Arabidopsis thaliana*-*Arabidopsis lyrata* subsp. *petraea* as an alternative model system for the study of polyploidy in plants." *Molecular genetics and genomics : MGG* (2009). 281(4): 421-435.
- Becker, C., J. Hagmann, J. Muller, D. Koenig, O. Stegle, K. Borgwardt and D. Weigel "Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome." *Nature* (2011). 480(7376): 245-249.
- Bender, J. and G. R. Fink "Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of *Arabidopsis*." *Cell* (1995). 83(5): 725-734.
- Berger, F., T. M. Vu, J. Li and B. Chen "Hypothesis: selection of imprinted genes is driven by silencing deleterious gene activity in somatic tissues." *Cold Spring Harbor symposia on quantitative biology* (2012). 77: 23-29.



- Berr, A., E. J. McCallum, R. Menard, D. Meyer, J. Fuchs, A. Dong and W. H. Shen "Arabidopsis SET DOMAIN GROUP2 is required for H3K4 trimethylation and is crucial for both sporophyte and gametophyte development." *The Plant cell* (2010). 22(10): 3232-3248.
- Berr, A., S. Shafiq and W. H. Shen "Histone modifications in transcriptional activation during plant development." *Biochimica et biophysica acta* (2011). 1809(10): 567-576.
- Bertrand, C., M. Benhamed, Y. F. Li, M. Ayadi, G. Lemonnier, J. P. Renou, M. Delarue and D. X. Zhou "Arabidopsis HAF2 gene encoding TATA-binding protein (TBP)-associated factor TAF1, is required to integrate light signals to regulate gene expression and growth." *The Journal of biological chemistry* (2005). 280(2): 1465-1473.
- Bertrand, C., C. Bergounioux, S. Domenichini, M. Delarue and D. X. Zhou "Arabidopsis histone acetyltransferase AtGCN5 regulates the floral meristem activity through the WUSCHEL/AGAMOUS pathway." *The Journal of biological chemistry* (2003). 278(30): 28246-28251.
- Bologna, N. G. and O. Voinnet "The diversity, biogenesis, and activities of endogenous silencing small RNAs in Arabidopsis." *Annu Rev Plant Biol* (2014). 65: 473-503.
- Bomblies, K., L. Yant, R. A. Laitinen, S. T. Kim, J. D. Hollister, N. Warthmann, J. Fitz and D. Weigel "Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of Arabidopsis thaliana." *PLoS Genet* (2010a). 6(3): e1000890.
- Bomblies, K., L. Yant, R. A. Laitinen, S. T. Kim, J. D. Hollister, N. Warthmann, J. Fitz and D. Weigel "Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of Arabidopsis thaliana." *PLoS genetics* (2010b). 6(3): e1000890.
- Bowman, J. L., D. R. Smyth and E. M. Meyerowitz "Genes directing flower development in Arabidopsis." *The Plant cell* (1989). 1(1): 37-52.
- Bratzel, F., G. Lopez-Torrejon, M. Koch, J. C. Del Pozo and M. Calonje "Keeping cell identity in Arabidopsis requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination." *Current biology : CB* (2010). 20(20): 1853-1859.
- Brzeska, K., J. Brzeski, J. Smith and V. L. Chandler "Transgenic expression of CBBP, a CXC domain protein, establishes paramutation in maize." *Proceedings of the National Academy of Sciences of the United States of America* (2010). 107(12): 5516-5521.
- Bushell, C., M. Spielman and R. J. Scott "The basis of natural and artificial postzygotic hybridization barriers in Arabidopsis species." *The Plant cell* (2003). 15(6): 1430-1442.
- Calarco, J. P., F. Borges, M. T. Donoghue, F. Van Ex, P. E. Jullien, T. Lopes, R. Gardner, F. Berger, J. A. Feijo, J. D. Becker and R. A. Martienssen "Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA." *Cell* (2012). 151(1): 194-205.
- Chen, M., M. Ha, E. Lackey, J. Wang and Z. J. Chen "RNAi of met1 reduces DNA methylation and induces genome-specific changes in gene expression and centromeric small RNA accumulation in Arabidopsis allopolyploids." *Genetics* (2008). 178(4): 1845-1858.





- Chodavarapu, R. K., S. Feng, Y. V. Bernatavichute, P. Y. Chen, H. Stroud, Y. Yu, J. A. Hetzel, F. Kuo, J. Kim, S. J. Cokus, D. Casero, M. Bernal, P. Huijser, A. T. Clark, U. Kramer, S. S. Merchant, X. Zhang, S. E. Jacobsen and M. Pellegrini "Relationship between nucleosome positioning and DNA methylation." *Nature* (2010). 466(7304): 388-392.
- Choi, Y., M. Gehring, L. Johnson, M. Hannon, J. J. Harada, R. B. Goldberg, S. E. Jacobsen and R. L. Fischer "DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis." *Cell* (2002). 110(1): 33-42.
- Coe, E. H. "A Regular and Continuing Conversion-Type Phenomenon at the B Locus in Maize." *Proceedings of the National Academy of Sciences of the United States of America* (1959). 45(6): 828-832.
- Cokus, S. J., S. Feng, X. Zhang, Z. Chen, B. Merriman, C. D. Haudenschild, S. Pradhan, S. F. Nelson, M. Pellegrini and S. E. Jacobsen "Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning." *Nature* (2008). 452(7184): 215-219.
- Coleman-Derr, D. and D. Zilberman "Deposition of histone variant H2A.Z within gene bodies regulates responsive genes." *PLoS Genet* (2012a). 8(10): e1002988.
- Coleman-Derr, D. and D. Zilberman "Deposition of histone variant H2A.Z within gene bodies regulates responsive genes." *PLoS genetics* (2012b). 8(10): e1002988.
- Comai, L., A. P. Tyagi, K. Winter, R. Holmes-Davis, S. H. Reynolds, Y. Stevens and B. Byers "Phenotypic instability and rapid gene silencing in newly formed arabidopsis allotetraploids." *The Plant cell* (2000). 12(9): 1551-1568.
- Coustham, V., D. Vlad, A. Deremetz, I. Gy, F. A. Cubillos, E. Kerdaffrec, O. Loudet and N. Bouche "SHOOT GROWTH1 maintains Arabidopsis epigenomes by regulating IBM1." *PloS one* (2014). 9(1): e84687.
- Creasey, K. M., J. Zhai, F. Borges, F. Van Ex, M. Regulski, B. C. Meyers and R. A. Martienssen "miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis." *Nature* (2014). 508(7496): 411-415.
- Cubas, P., C. Vincent and E. Coen "An epigenetic mutation responsible for natural variation in floral symmetry." *Nature* (1999). 401(6749): 157-161.
- Cui, H. and N. V. Fedoroff "Inducible DNA demethylation mediated by the maize Suppressor-mutator transposon-encoded TnpA protein." *Plant Cell* (2002). 14(11): 2883-2899.
- Deleris, A., H. Stroud, Y. Bernatavichute, E. Johnson, G. Klein, D. Schubert and S. E. Jacobsen "Loss of the DNA methyltransferase MET1 Induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in Arabidopsis thaliana." *PLoS Genet* (2012). 8(11): e1003062.
- Dhawan, R., H. Luo, A. M. Foerster, S. Abuqamar, H. N. Du, S. D. Briggs, O. Mittelsten Scheid and T. Mengiste "HISTONE MONOUBIQUITINATION1 interacts with a subunit of the mediator complex and regulates defense against necrotrophic fungal pathogens in Arabidopsis." *The Plant cell* (2009). 21(3): 1000-1019.
- Downen, R. H., M. Pelizzola, R. J. Schmitz, R. Lister, J. M. Downen, J. R. Nery, J. E. Dixon and J. R. Ecker "Widespread dynamic DNA methylation in response to biotic stress." *Proc Natl Acad Sci U S A* (2012a). 109(32): E2183-2191.



- Downen, R. H., M. Pelizzola, R. J. Schmitz, R. Lister, J. M. Downen, J. R. Nery, J. E. Dixon and J. R. Ecker "Widespread dynamic DNA methylation in response to biotic stress." *Proceedings of the National Academy of Sciences of the United States of America* (2012b). 109(32): E2183-2191.
- Du, J., L. M. Johnson, M. Groth, S. Feng, C. J. Hale, S. Li, A. A. Vashisht, J. Gallego-Bartolome, J. A. Wohlschlegel, D. J. Patel and S. E. Jacobsen "Mechanism of DNA methylation-directed histone methylation by KRYPTONITE." *Mol Cell* (2014). 55(3): 495-504.
- Du, J., X. Zhong, Y. V. Bernatavichute, H. Stroud, S. Feng, E. Caro, A. A. Vashisht, J. Terragni, H. G. Chin, A. Tu, J. Hetzel, J. A. Wohlschlegel, S. Pradhan, D. J. Patel and S. E. Jacobsen "Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants." *Cell* (2012). 151(1): 167-180.
- Durand, S., N. Bouche, E. Perez Strand, O. Loudet and C. Camilleri "Rapid establishment of genetic incompatibility through natural epigenetic variation." *Current biology : CB* (2012). 22(4): 326-331.
- Earley, K., R. J. Lawrence, O. Pontes, R. Reuther, A. J. Enciso, M. Silva, N. Neves, M. Gross, W. Viegas and C. S. Pikaard "Erasure of histone acetylation by Arabidopsis HDA6 mediates large-scale gene silencing in nucleolar dominance." *Genes & development* (2006). 20(10): 1283-1293.
- Earley, K. W., F. Pontvianne, A. T. Wierzbicki, T. Blevins, S. Tucker, P. Costa-Nunes, O. Pontes and C. S. Pikaard "Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation." *Genes & development* (2010). 24(11): 1119-1132.
- Eun, C., Z. J. Lorkovic, T. Sasaki, U. Naumann, A. J. Matzke and M. Matzke "Use of forward genetic screens to identify genes required for RNA-directed DNA methylation in *Arabidopsis thaliana*." *Cold Spring Harb Symp Quant Biol* (2012). 77: 195-204.
- Feng, S., S. J. Cokus, X. Zhang, P. Y. Chen, M. Bostick, M. G. Goll, J. Hetzel, J. Jain, S. H. Strauss, M. E. Halpern, C. Ukomadu, K. C. Sadler, S. Pradhan, M. Pellegrini and S. E. Jacobsen "Conservation and divergence of methylation patterning in plants and animals." *Proceedings of the National Academy of Sciences of the United States of America* (2010). 107(19): 8689-8694.
- FitzGerald, J., M. Luo, A. Chaudhury and F. Berger "DNA methylation causes predominant maternal controls of plant embryo growth." *PloS one* (2008). 3(5): e2298.
- Fu, Y., A. Kawabe, M. Etcheverry, T. Ito, A. Toyoda, A. Fujiyama, V. Colot, Y. Tarutani and T. Kakutani "Mobilization of a plant transposon by expression of the transposon-encoded anti-silencing factor." *EMBO J* (2013). 32(17): 2407-2417.
- Gao, Z., H. L. Liu, L. Daxinger, O. Pontes, X. He, W. Qian, H. Lin, M. Xie, Z. J. Lorkovic, S. Zhang, D. Miki, X. Zhan, D. Pontier, T. Lagrange, H. Jin, A. J. Matzke, M. Matzke, C. S. Pikaard and J. K. Zhu "An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation." *Nature* (2010). 465(7294): 106-109.
- Gehring, M., J. H. Huh, T. F. Hsieh, J. Penterman, Y. Choi, J. J. Harada, R. B. Goldberg and R. L. Fischer "DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation." *Cell* (2006). 124(3): 495-506.
- Gehring, M., V. Missirlian and S. Henikoff "Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds." *PloS one* (2011). 6(8): e23687.



- Gong, Z., T. Morales-Ruiz, R. R. Ariza, T. Roldan-Arjona, L. David and J. K. Zhu "ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase." *Cell* (2002). 111(6): 803-814.
- Greaves, I. K., M. Groszmann, H. Ying, J. M. Taylor, W. J. Peacock and E. S. Dennis "Trans chromosomal methylation in Arabidopsis hybrids." *Proc Natl Acad Sci U S A* (2012a). 109(9): 3570-3575.
- Greaves, I. K., M. Groszmann, H. Ying, J. M. Taylor, W. J. Peacock and E. S. Dennis "Trans chromosomal methylation in Arabidopsis hybrids." *Proceedings of the National Academy of Sciences of the United States of America* (2012b). 109(9): 3570-3575.
- Greenberg, M. V., A. Deleris, C. J. Hale, A. Liu, S. Feng and S. E. Jacobsen "Interplay between active chromatin marks and RNA-directed DNA methylation in Arabidopsis thaliana." *PLoS genetics* (2013). 9(11): e1003946.
- Grini, P. E., T. Thorstensen, V. Alm, G. Vizcay-Barrena, S. S. Windju, T. S. Jorstad, Z. A. Wilson and R. B. Aalen "The ASH1 HOMOLOG 2 (ASHH2) histone H3 methyltransferase is required for ovule and anther development in Arabidopsis." *PloS one* (2009). 4(11): e7817.
- Groszmann, M., I. K. Greaves, Z. I. Albertyn, G. N. Scofield, W. J. Peacock and E. S. Dennis "Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor." *Proc Natl Acad Sci U S A* (2011a). 108(6): 2617-2622.
- Groszmann, M., I. K. Greaves, Z. I. Albertyn, G. N. Scofield, W. J. Peacock and E. S. Dennis "Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor." *Proceedings of the National Academy of Sciences of the United States of America* (2011b). 108(6): 2617-2622.
- Groth, M., H. Stroud, S. Feng, M. V. Greenberg, A. A. Vashisht, J. A. Wohlschlegel, S. E. Jacobsen and I. Ausin "SNF2 chromatin remodeler-family proteins FRG1 and -2 are required for RNA-directed DNA methylation." *Proc Natl Acad Sci U S A* (2014). 111(49): 17666-17671.
- Gu, X., D. Jiang, Y. Wang, A. Bachmair and Y. He "Repression of the floral transition via histone H2B monoubiquitination." *The Plant journal : for cell and molecular biology* (2009). 57(3): 522-533.
- Guo, W., P. Fiziev, W. Yan, S. Cokus, X. Sun, M. Q. Zhang, P. Y. Chen and M. Pellegrini "BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data." *BMC genomics* (2013). 14: 774.
- Haag, J. R. and C. S. Pikaard "Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing." *Nature reviews. Molecular cell biology* (2011). 12(8): 483-492.
- Hagmann, J., C. Becker, J. Muller, O. Stegle, R. C. Meyer, G. Wang, K. Schneeberger, J. Fitz, T. Altmann, J. Bergelson, K. Borgwardt and D. Weigel "Century-scale methylome stability in a recently diverged Arabidopsis thaliana lineage." *PLoS Genet* (2015). 11(1): e1004920.
- Hajheidari, M., C. Koncz and D. Eick "Emerging roles for RNA polymerase II CTD in Arabidopsis." *Trends in plant science* (2013). 18(11): 633-643.



- Hale, C. J., J. L. Stonaker, S. M. Gross and J. B. Hollick "A novel Snf2 protein maintains trans-generational regulatory states established by paramutation in maize." *PLoS biology* (2007). 5(10): e275.
- Havecker, E. R., L. M. Wallbridge, P. Fedito, T. J. Hardcastle and D. C. Baulcombe "Metastable differentially methylated regions within Arabidopsis inbred populations are associated with modified expression of non-coding transcripts." *PloS one* (2012). 7(9): e45242.
- He, C., H. Huang and L. Xu "Mechanisms guiding Polycomb activities during gene silencing in Arabidopsis thaliana." *Frontiers in plant science* (2013). 4: 454.
- He, G., X. Zhu, A. A. Elling, L. Chen, X. Wang, L. Guo, M. Liang, H. He, H. Zhang, F. Chen, Y. Qi, R. Chen and X. W. Deng "Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids." *The Plant cell* (2010). 22(1): 17-33.
- Henderson, I. R. and S. E. Jacobsen "Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading." *Genes Dev* (2008). 22(12): 1597-1606.
- Hoppmann, V., T. Thorstensen, P. E. Kristiansen, S. V. Veiseth, M. A. Rahman, K. Finne, R. B. Aalen and R. Aasland "The CW domain, a new histone recognition module in chromatin proteins." *The EMBO journal* (2011). 30(10): 1939-1952.
- Hsieh, T. F., J. Shin, R. Uzawa, P. Silva, S. Cohen, M. J. Bauer, M. Hashimoto, R. C. Kirkbride, J. J. Harada, D. Zilberman and R. L. Fischer "Regulation of imprinted gene expression in Arabidopsis endosperm." *Proceedings of the National Academy of Sciences of the United States of America* (2011). 108(5): 1755-1762.
- Huang, C. F., D. Miki, K. Tang, H. R. Zhou, Z. Zheng, W. Chen, Z. Y. Ma, L. Yang, H. Zhang, R. Liu, X. J. He and J. K. Zhu "A Pre-mRNA-splicing factor is required for RNA-directed DNA methylation in Arabidopsis." *PLoS Genet* (2013). 9(9): e1003779.
- Huettel, B., T. Kanno, L. Daxinger, W. Aufsatz, A. J. Matzke and M. Matzke "Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis." *The EMBO journal* (2006). 25(12): 2828-2836.
- Inagaki, S., A. Miura-Kamio, Y. Nakamura, F. Lu, X. Cui, X. Cao, H. Kimura, H. Saze and T. Kakutani "Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome." *The EMBO journal* (2010a). 29(20): 3496-3506.
- Inagaki, S., A. Miura-Kamio, Y. Nakamura, F. Lu, X. Cui, X. Cao, H. Kimura, H. Saze and T. Kakutani "Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome." *EMBO J* (2010b). 29(20): 3496-3506.
- Ingouff, M., S. Rademacher, S. Holec, L. Soljic, N. Xin, A. Readshaw, S. H. Foo, B. Lahouze, S. Sprunck and F. Berger "Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in Arabidopsis." *Current biology : CB* (2010). 20(23): 2137-2143.
- Irizarry, R. A., C. Ladd-Acosta, B. Carvalho, H. Wu, S. A. Brandenburg, J. A. Jeddloh, B. Wen and A. P. Feinberg "Comprehensive high-throughput arrays for relative methylation (CHARM)." *Genome Res* (2008). 18(5): 780-790.
- Jacobsen, S. E. and E. M. Meyerowitz "Hypermethylated SUPERMAN epigenetic alleles in arabidopsis." *Science* (1997). 277(5329): 1100-1103.





- Jacobsen, S. E., H. Sakai, E. J. Finnegan, X. Cao and E. M. Meyerowitz "Ectopic hypermethylation of flower-specific genes in Arabidopsis." *Curr Biol* (2000a). 10(4): 179-186.
- Jacobsen, S. E., H. Sakai, E. J. Finnegan, X. Cao and E. M. Meyerowitz "Ectopic hypermethylation of flower-specific genes in Arabidopsis." *Current biology : CB* (2000b). 10(4): 179-186.
- Jiang, D., X. Gu and Y. He "Establishment of the winter-annual growth habit via FRIGIDA-mediated histone methylation at FLOWERING LOCUS C in Arabidopsis." *The Plant cell* (2009). 21(6): 1733-1746.
- Jiang, D., N. C. Kong, X. Gu, Z. Li and Y. He "Arabidopsis COMPASS-like complexes mediate histone H3 lysine-4 trimethylation to control floral transition and plant development." *PLoS genetics* (2011). 7(3): e1001330.
- Jiang, D., Y. Wang and Y. He "Repression of FLOWERING LOCUS C and FLOWERING LOCUS T by the Arabidopsis Polycomb repressive complex 2 components." *PloS one* (2008). 3(10): e3404.
- Jin, H., W. Hu, Z. Wei, L. Wan, G. Li, G. Tan, L. Zhu and G. He "Alterations in cytosine methylation and species-specific transcription induced by interspecific hybridization between *Oryza sativa* and *O. officinalis*." *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* (2008). 117(8): 1271-1279.
- Johannes, F., E. Porcher, F. K. Teixeira, V. Saliba-Colombani, M. Simon, N. Agier, A. Bulski, J. Albuissou, F. Heredia, P. Audigier, D. Bouchez, C. Dillmann, P. Guerche, F. Hospital and V. Colot "Assessing the impact of transgenerational epigenetic variation on complex traits." *PLoS Genet* (2009a). 5(6): e1000530.
- Johannes, F., E. Porcher, F. K. Teixeira, V. Saliba-Colombani, M. Simon, N. Agier, A. Bulski, J. Albuissou, F. Heredia, P. Audigier, D. Bouchez, C. Dillmann, P. Guerche, F. Hospital and V. Colot "Assessing the impact of transgenerational epigenetic variation on complex traits." *PLoS genetics* (2009b). 5(6): e1000530.
- Josefsson, C., B. Dilkes and L. Comai "Parent-dependent loss of gene silencing during interspecies hybridization." *Curr Biol* (2006a). 16(13): 1322-1328.
- Josefsson, C., B. Dilkes and L. Comai "Parent-dependent loss of gene silencing during interspecies hybridization." *Current biology : CB* (2006b). 16(13): 1322-1328.
- Jullien, P. E., T. Kinoshita, N. Ohad and F. Berger "Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting." *The Plant cell* (2006). 18(6): 1360-1372.
- Jullien, P. E., A. Mosquna, M. Ingouff, T. Sakata, N. Ohad and F. Berger "Retinoblastoma and its binding partner MSI1 control imprinting in Arabidopsis." *PLoS biology* (2008). 6(8): e194.
- Kakutani, T. "Genetic characterization of late-flowering traits induced by DNA hypomethylation mutation in Arabidopsis thaliana." *The Plant journal : for cell and molecular biology* (1997). 12(6): 1447-1451.
- Kakutani, T., K. Munakata, E. J. Richards and H. Hirochika "Meiotically and mitotically stable inheritance of DNA hypomethylation induced by *ddm1* mutation of Arabidopsis thaliana." *Genetics* (1999). 151(2): 831-838.



- Kankel, M. W., D. E. Ramsey, T. L. Stokes, S. K. Flowers, J. R. Haag, J. A. Jeddelloh, N. C. Riddle, M. L. Verbsky and E. J. Richards "Arabidopsis MET1 cytosine methyltransferase mutants." *Genetics* (2003). 163(3): 1109-1122.
- Kanno, T., W. Aufsatz, E. Jaligot, M. F. Mette, M. Matzke and A. J. Matzke "A SNF2-like protein facilitates dynamic control of DNA methylation." *EMBO reports* (2005). 6(7): 649-655.
- Kanno, T. and Y. Habu "siRNA-mediated chromatin maintenance and its function in *Arabidopsis thaliana*." *Biochim Biophys Acta* (2011). 1809(8): 444-451.
- Kapoor, A., F. Agius and J. K. Zhu "Preventing transcriptional gene silencing by active DNA demethylation." *FEBS letters* (2005). 579(26): 5889-5898.
- Kato, M., A. Miura, J. Bender, S. E. Jacobsen and T. Kakutani "Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*." *Curr Biol* (2003a). 13(5): 421-426.
- Kato, M., A. Miura, J. Bender, S. E. Jacobsen and T. Kakutani "Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*." *Current biology : CB* (2003b). 13(5): 421-426.
- Kato, M., K. Takashima and T. Kakutani "Epigenetic control of CACTA transposon mobility in *Arabidopsis thaliana*." *Genetics* (2004). 168(2): 961-969.
- Keshet, I., Y. Schlesinger, S. Farkash, E. Rand, M. Hecht, E. Segal, E. Pikarski, R. A. Young, A. Niveleau, H. Cedar and I. Simon "Evidence for an instructive mechanism of de novo methylation in cancer cells." *Nat Genet* (2006). 38(2): 149-153.
- Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley and S. L. Salzberg "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions." *Genome Biol* (2013). 14(4): R36.
- Kinoshita, T., A. Miura, Y. Choi, Y. Kinoshita, X. Cao, S. E. Jacobsen, R. L. Fischer and T. Kakutani "One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation." *Science* (2004). 303(5657): 521-523.
- Kinoshita, T., R. Yadegari, J. J. Harada, R. B. Goldberg and R. L. Fischer "Imprinting of the MEDEA polycomb gene in the *Arabidopsis* endosperm." *The Plant cell* (1999). 11(10): 1945-1952.
- Langmead, B. and S. L. Salzberg "Fast gapped-read alignment with Bowtie 2." *Nat Methods* (2012). 9(4): 357-359.
- Law, J. A. and S. E. Jacobsen "Establishing, maintaining and modifying DNA methylation patterns in plants and animals." *Nat Rev Genet* (2010). 11(3): 204-220.
- Law, J. A., A. A. Vashisht, J. A. Wohlschlegel and S. E. Jacobsen "SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV." *PLoS Genet* (2011). 7(7): e1002195.
- Lee, H. S. and Z. J. Chen "Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids." *Proceedings of the National Academy of Sciences of the United States of America* (2001). 98(12): 6753-6758.



- Lei, M., H. La, K. Lu, P. Wang, D. Miki, Z. Ren, C. G. Duan, X. Wang, K. Tang, L. Zeng, L. Yang, H. Zhang, W. Nie, P. Liu, J. Zhou, R. Liu, Y. Zhong, D. Liu and J. K. Zhu "Arabidopsis EDM2 promotes IBM1 distal polyadenylation and regulates genome DNA methylation patterns." *Proceedings of the National Academy of Sciences of the United States of America* (2014). 111(1): 527-532.
- Li, H. and R. Durbin "Fast and accurate short read alignment with Burrows-Wheeler transform." *Bioinformatics* (2009). 25(14): 1754-1760.
- Li, X., W. Qian, Y. Zhao, C. Wang, J. Shen, J. K. Zhu and Z. Gong "Antisilencing role of the RNA-directed DNA methylation pathway and a histone acetyltransferase in Arabidopsis." *Proceedings of the National Academy of Sciences of the United States of America* (2012). 109(28): 11425-11430.
- Lindroth, A. M., X. Cao, J. P. Jackson, D. Zilberman, C. M. McCallum, S. Henikoff and S. E. Jacobsen "Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation." *Science* (2001). 292(5524): 2077-2080.
- Lippman, Z., B. May, C. Yordan, T. Singer and R. Martienssen "Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification." *PLoS biology* (2003a). 1(3): E67.
- Lippman, Z., B. May, C. Yordan, T. Singer and R. Martienssen "Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification." *PLoS Biol* (2003b). 1(3): E67.
- Lisanti, S., T. von Zglinicki and J. C. Mathers "Standardization and quality controls for the methylated DNA immunoprecipitation technique." *Epigenetics* (2012). 7(6): 615-625.
- Lister, R., R. C. O'Malley, J. Tonti-Filippini, B. D. Gregory, C. C. Berry, A. H. Millar and J. R. Ecker "Highly integrated single-base resolution maps of the epigenome in Arabidopsis." *Cell* (2008). 133(3): 523-536.
- Liu, C., F. Lu, X. Cui and X. Cao "Histone methylation in higher plants." *Annual review of plant biology* (2010). 61: 395-420.
- Lolas, I. B., K. Himanen, J. T. Gronlund, C. Lynggaard, A. Houben, M. Melzer, M. Van Lijsebettens and K. D. Grasser "The transcript elongation factor FACT affects Arabidopsis vegetative and reproductive development and genetically interacts with HUB1/2." *The Plant journal : for cell and molecular biology* (2010). 61(4): 686-697.
- Lu, F., X. Cui, S. Zhang, T. Jenuwein and X. Cao "Arabidopsis REF6 is a histone H3 lysine 27 demethylase." *Nature genetics* (2011). 43(7): 715-719.
- Lu, F., X. Cui, S. Zhang, C. Liu and X. Cao "JMJ14 is an H3K4 demethylase regulating flowering time in Arabidopsis." *Cell research* (2010). 20(3): 387-390.
- Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond "Crystal structure of the nucleosome core particle at 2.8 Å resolution." *Nature* (1997). 389(6648): 251-260.
- Luo, M., P. Bilodeau, E. S. Dennis, W. J. Peacock and A. Chaudhury "Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds." *Proceedings of the National Academy of Sciences of the United States of America* (2000). 97(19): 10637-10642.



- Madlung, A., R. W. Masuelli, B. Watson, S. H. Reynolds, J. Davison and L. Comai "Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic Arabidopsis allotetraploids." *Plant physiology* (2002). 129(2): 733-746.
- Madlung, A., A. P. Tyagi, B. Watson, H. Jiang, T. Kagochi, R. W. Doerge, R. Martienssen and L. Comai "Genomic changes in synthetic Arabidopsis polyploids." *The Plant journal : for cell and molecular biology* (2005). 41(2): 221-230.
- Magoc, T. and S. L. Salzberg "FLASH: fast length adjustment of short reads to improve genome assemblies." *Bioinformatics* (2011). 27(21): 2957-2963.
- Martinez-Macias, M. I., D. Cordoba-Canero, R. R. Ariza and T. Roldan-Arjona "The DNA repair protein XRCC1 functions in the plant DNA demethylation pathway by stimulating cytosine methylation (5-meC) excision, gap tailoring, and DNA ligation." *The Journal of biological chemistry* (2013). 288(8): 5496-5505.
- Martinez-Macias, M. I., W. Qian, D. Miki, O. Pontes, Y. Liu, K. Tang, R. Liu, T. Morales-Ruiz, R. R. Ariza, T. Roldan-Arjona and J. K. Zhu "A DNA 3' phosphatase functions in active DNA demethylation in Arabidopsis." *Molecular cell* (2012). 45(3): 357-370.
- Mathieu, O., A. V. Probst and J. Paszkowski "Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis." *EMBO J* (2005a). 24(15): 2783-2791.
- Mathieu, O., A. V. Probst and J. Paszkowski "Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis." *The EMBO journal* (2005b). 24(15): 2783-2791.
- Mathieu, O., J. Reinders, M. Caikovski, C. Smathajitt and J. Paszkowski "Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation." *Cell* (2007). 130(5): 851-862.
- Matsushita, Y., M. Deguchi, M. Youda, M. Nishiguchi and H. Nyunoya "The tomato mosaic tobamovirus movement protein interacts with a putative transcriptional coactivator KELP." *Mol Cells* (2001). 12(1): 57-66.
- McKeown, P. C., S. Laouielle-Duprat, P. Prins, P. Wolff, M. W. Schmid, M. T. Donoghue, A. Fort, D. Duszynska, A. Comte, N. T. Lao, T. J. Wennblom, G. Smant, C. Kohler, U. Grossniklaus and C. Spillane "Identification of imprinted genes subject to parent-of-origin specific expression in Arabidopsis thaliana seeds." *BMC plant biology* (2011). 11: 113.
- Mirouze, M., J. Reinders, E. Bucher, T. Nishimura, K. Schneeberger, S. Ossowski, J. Cao, D. Weigel, J. Paszkowski and O. Mathieu "Selective epigenetic control of retrotransposition in Arabidopsis." *Nature* (2009). 461(7262): 427-430.
- Miura, A., M. Nakamura, S. Inagaki, A. Kobayashi, H. Saze and T. Kakutani "An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites." *The EMBO journal* (2009). 28(8): 1078-1086.
- Miura, A., S. Yonebayashi, K. Watanabe, T. Toyama, H. Shimada and T. Kakutani "Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis." *Nature* (2001). 411(6834): 212-214.
- Morales-Ruiz, T., A. P. Ortega-Galisteo, M. I. Ponferrada-Marin, M. I. Martinez-Macias, R. R. Ariza and T. Roldan-Arjona "DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases." *Proceedings of the National Academy of Sciences of the United States of America* (2006). 103(18): 6853-6858.





- Nuthikattu, S., A. D. McCue, K. Panda, D. Fultz, C. DeFraia, E. N. Thomas and R. K. Slotkin "The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs." *Plant Physiol* (2013). 162(1): 116-131.
- O'Neill, R. J., M. J. O'Neill and J. A. Graves "Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid." *Nature* (1998). 393(6680): 68-72.
- Ortega-Galisteo, A. P., T. Morales-Ruiz, R. R. Ariza and T. Roldan-Arjona "Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks." *Plant molecular biology* (2008). 67(6): 671-681.
- Ossowski, S., K. Schneeberger, R. M. Clark, C. Lanz, N. Warthmann and D. Weigel "Sequencing of natural strains of *Arabidopsis thaliana* with short reads." *Genome Res* (2008). 18(12): 2024-2033.
- Pandey, R., A. Muller, C. A. Napoli, D. A. Selinger, C. S. Pikaard, E. J. Richards, J. Bender, D. W. Mount and R. A. Jorgensen "Analysis of histone acetyltransferase and histone deacetylase families of *Arabidopsis thaliana* suggests functional diversification of chromatin modification among multicellular eukaryotes." *Nucleic acids research* (2002). 30(23): 5036-5055.
- Penterman, J., R. Uzawa and R. L. Fischer "Genetic interactions between DNA demethylation and methylation in *Arabidopsis*." *Plant physiology* (2007a). 145(4): 1549-1557.
- Penterman, J., D. Zilberman, J. H. Huh, T. Ballinger, S. Henikoff and R. L. Fischer "DNA demethylation in the *Arabidopsis* genome." *Proceedings of the National Academy of Sciences of the United States of America* (2007b). 104(16): 6752-6757.
- Pickart, C. M. "Mechanisms underlying ubiquitination." *Annual review of biochemistry* (2001). 70: 503-533.
- Pignatta, D., R. M. Erdmann, E. Scheer, C. L. Picard, G. W. Bell and M. Gehring "Natural epigenetic polymorphisms lead to intraspecific variation in *Arabidopsis* gene imprinting." *eLife* (2014). e03198.
- Qian, W., D. Miki, H. Zhang, Y. Liu, X. Zhang, K. Tang, Y. Kan, H. La, X. Li, S. Li, X. Zhu, X. Shi, K. Zhang, O. Pontes, X. Chen, R. Liu, Z. Gong and J. K. Zhu "A histone acetyltransferase regulates active DNA demethylation in *Arabidopsis*." *Science* (2012). 336(6087): 1445-1448.
- Raissig, M. T., M. Bemer, C. Baroux and U. Grossniklaus "Genomic imprinting in the *Arabidopsis* embryo is partly regulated by PRC2." *PLoS genetics* (2013). 9(12): e1003862.
- Rangwala, S. H., R. Elumalai, C. Vanier, H. Ozkan, D. W. Galbraith and E. J. Richards "Meiotically stable natural epialleles of *Sadhu*, a novel *Arabidopsis* retroposon." *PLoS genetics* (2006). 2(3): e36.
- Redon, C., D. Pilch, E. Rogakou, O. Sedelnikova, K. Newrock and W. Bonner "Histone H2A variants H2AX and H2AZ." *Current opinion in genetics & development* (2002). 12(2): 162-169.



- Reinders, J., B. B. Wulff, M. Mirouze, A. Mari-Ordonez, M. Dapp, W. Rozhon, E. Bucher, G. Theiler and J. Paszkowski "Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes." *Genes & development* (2009a). 23(8): 939-950.
- Reinders, J., B. B. Wulff, M. Mirouze, A. Mari-Ordonez, M. Dapp, W. Rozhon, E. Bucher, G. Theiler and J. Paszkowski "Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes." *Genes Dev* (2009b). 23(8): 939-950.
- Richards, E. J. "Quantitative epigenetics: DNA sequence variation need not apply." *Genes & development* (2009). 23(14): 1601-1605.
- Rigal, M., Z. Kevei, T. Pelissier and O. Mathieu "DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns." *EMBO J* (2012). 31(13): 2981-2993.
- Rigal, M. and O. Mathieu "A "mille-feuille" of silencing: epigenetic control of transposable elements." *Biochim Biophys Acta* (2011). 1809(8): 452-458.
- Roudier, F., I. Ahmed, C. Berard, A. Sarazin, T. Mary-Huard, S. Cortijo, D. Bouyer, E. Caillieux, E. Duvernois-Berthet, L. Al-Shikhley, L. Giraut, B. Despres, S. Drevensek, F. Barneche, S. Derozier, V. Brunaud, S. Aubourg, A. Schnittger, C. Bowler, M. L. Martin-Magniette, S. Robin, M. Caboche and V. Colot "Integrative epigenomic mapping defines four main chromatin states in Arabidopsis." *The EMBO journal* (2011). 30(10): 1928-1938.
- Sasaki, T., A. Kobayashi, H. Saze and T. Kakutani "RNAi-independent de novo DNA methylation revealed in Arabidopsis mutants of chromatin remodeling gene DDM1." *Plant J* (2012). 70(5): 750-758.
- Saze, H. and T. Kakutani "Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1." *The EMBO journal* (2007a). 26(15): 3641-3652.
- Saze, H. and T. Kakutani "Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1." *EMBO J* (2007b). 26(15): 3641-3652.
- Saze, H., J. Kitayama, K. Takashima, S. Miura, Y. Harukawa, T. Ito and T. Kakutani "Mechanism for full-length RNA processing of Arabidopsis genes containing intragenic heterochromatin." *Nature communications* (2013a). 4: 2301.
- Saze, H., J. Kitayama, K. Takashima, S. Miura, Y. Harukawa, T. Ito and T. Kakutani "Mechanism for full-length RNA processing of Arabidopsis genes containing intragenic heterochromatin." *Nat Commun* (2013b). 4: 2301.
- Saze, H., O. Mittelsten Scheid and J. Paszkowski "Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis." *Nat Genet* (2003). 34(1): 65-69.
- Saze, H., A. Shiraishi, A. Miura and T. Kakutani "Control of genic DNA methylation by a jmjC domain-containing protein in Arabidopsis thaliana." *Science* (2008). 319(5862): 462-465.
- Schlappi, M., R. Raina and N. Fedoroff "Epigenetic regulation of the maize Spm transposable element: novel activation of a methylated promoter by TnpA." *Cell* (1994). 77(3): 427-437.



- Schmitz, R. J., M. D. Schultz, M. G. Lewsey, R. C. O'Malley, M. A. Urich, O. Libiger, N. J. Schork and J. R. Ecker "Transgenerational epigenetic instability is a source of novel methylation variants." *Science* (2011). 334(6054): 369-373.
- Schmitz, R. J., M. D. Schultz, M. A. Urich, J. R. Nery, M. Pelizzola, O. Libiger, A. Alix, R. B. McCosh, H. Chen, N. J. Schork and J. R. Ecker "Patterns of population epigenomic diversity." *Nature* (2013). 495(7440): 193-198.
- Schmitz, R. J., Y. Tamada, M. R. Doyle, X. Zhang and R. M. Amasino "Histone H2B deubiquitination is required for transcriptional activation of FLOWERING LOCUS C and for proper control of flowering in Arabidopsis." *Plant physiology* (2009). 149(2): 1196-1204.
- Schwartz, Y. B. and V. Pirrotta "Polycomb complexes and epigenetic states." *Current opinion in cell biology* (2008). 20(3): 266-273.
- Sequeira-Mendes, J., I. Araguez, R. Peiro, R. Mendez-Giraldez, X. Zhang, S. E. Jacobsen, U. Bastolla and C. Gutierrez "The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States." *The Plant cell* (2014).
- Shen, H., H. He, J. Li, W. Chen, X. Wang, L. Guo, Z. Peng, G. He, S. Zhong, Y. Qi, W. Terzaghi and X. W. Deng "Genome-wide analysis of DNA methylation and gene expression changes in two Arabidopsis ecotypes and their reciprocal hybrids." *The Plant cell* (2012). 24(3): 875-892.
- Shu, H., M. Nakamura, A. Siretskiy, L. Borghi, I. Moraes, T. Wildhaber, W. Gruissem and L. Hennig "Arabidopsis replacement histone variant H3.3 occupies promoters of regulated genes." *Genome biology* (2014). 15(4): R62.
- Silveira, A. B., C. Trontin, S. Cortijo, J. Barau, L. E. Del Bem, O. Loudet, V. Colot and M. Vincentz "Extensive natural epigenetic variation at a de novo originated gene." *PLoS genetics* (2013). 9(4): e1003437.
- Singer, T., C. Yordan and R. A. Martienssen "Robertson's Mutator transposons in *A. thaliana* are regulated by the chromatin-remodeling gene *Decrease in DNA Methylation (DDM1)*." *Genes Dev* (2001). 15(5): 591-602.
- Slotkin, R. K., M. Vaughn, F. Borges, M. Tanurdzic, J. D. Becker, J. A. Feijo and R. A. Martienssen "Epigenetic reprogramming and small RNA silencing of transposable elements in pollen." *Cell* (2009). 136(3): 461-472.
- Soppe, W. J., S. E. Jacobsen, C. Alonso-Blanco, J. P. Jackson, T. Kakutani, M. Koornneef and A. J. Peeters "The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene." *Molecular cell* (2000a). 6(4): 791-802.
- Soppe, W. J., S. E. Jacobsen, C. Alonso-Blanco, J. P. Jackson, T. Kakutani, M. Koornneef and A. J. Peeters "The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene." *Mol Cell* (2000b). 6(4): 791-802.
- Sridhar, V. V., A. Kapoor, K. Zhang, J. Zhu, T. Zhou, P. M. Hasegawa, R. A. Bressan and J. K. Zhu "Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination." *Nature* (2007). 447(7145): 735-738.



- Steimer, A., P. Amedeo, K. Afsar, P. Frasz, O. Mittelsten Scheid and J. Paszkowski "Endogenous targets of transcriptional gene silencing in Arabidopsis." *Plant Cell* (2000). 12(7): 1165-1178.
- Stokes, T. L., B. N. Kunkel and E. J. Richards "Epigenetic variation in Arabidopsis disease resistance." *Genes Dev* (2002). 16(2): 171-182.
- Stroud, H., T. Do, J. Du, X. Zhong, S. Feng, L. Johnson, D. J. Patel and S. E. Jacobsen "Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis." *Nat Struct Mol Biol* (2014). 21(1): 64-72.
- Stroud, H., M. V. Greenberg, S. Feng, Y. V. Bernatavichute and S. E. Jacobsen "Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome." *Cell* (2013). 152(1-2): 352-364.
- Stroud, H., C. J. Hale, S. Feng, E. Caro, Y. Jacob, S. D. Michaels and S. E. Jacobsen "DNA methyltransferases are required to induce heterochromatic re-replication in Arabidopsis." *PLoS Genet* (2012). 8(7): e1002808.
- Tanurdzic, M., M. W. Vaughn, H. Jiang, T. J. Lee, R. K. Slotkin, B. Sosinski, W. F. Thompson, R. W. Doerge and R. A. Martienssen "Epigenomic consequences of immortalized plant cell suspension culture." *PLoS Biol* (2008). 6(12): 2880-2895.
- Tariq, M., H. Saze, A. V. Probst, J. Lichota, Y. Habu and J. Paszkowski "Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin." *Proceedings of the National Academy of Sciences of the United States of America* (2003). 100(15): 8823-8827.
- Teixeira, F. K., F. Heredia, A. Sarazin, F. Roudier, M. Boccara, C. Ciaudo, C. Cruaud, J. Poulain, M. Berdasco, M. F. Fraga, O. Voinnet, P. Wincker, M. Esteller and V. Colot "A role for RNAi in the selective correction of DNA methylation defects." *Science* (2009). 323(5921): 1600-1604.
- Tian, L. and Z. J. Chen "Blocking histone deacetylation in Arabidopsis induces pleiotropic effects on plant gene regulation and development." *Proceedings of the National Academy of Sciences of the United States of America* (2001). 98(1): 200-205.
- Tian, L., M. P. Fong, J. J. Wang, N. E. Wei, H. Jiang, R. W. Doerge and Z. J. Chen "Reversible histone acetylation and deacetylation mediate genome-wide, promoter-dependent and locus-specific changes in gene expression during plant development." *Genetics* (2005). 169(1): 337-345.
- Tian, L., J. Wang, M. P. Fong, M. Chen, H. Cao, S. B. Gelvin and Z. J. Chen "Genetic control of developmental changes induced by disruption of Arabidopsis histone deacetylase 1 (AtHD1) expression." *Genetics* (2003). 165(1): 399-409.
- Tittel-Elmer, M., E. Bucher, L. Broger, O. Mathieu, J. Paszkowski and I. Vaillant "Stress-induced activation of heterochromatic transcription." *PLoS Genet* (2010). 6(10): e1001175.
- Tomazou, E. M., V. K. Rakyar, G. Lefebvre, R. Andrews, P. Ellis, D. K. Jackson, C. Langford, M. D. Francis, L. Backdahl, M. Miretti, P. Coghill, D. Ottaviani, D. Sheer, A. Murrell and S. Beck "Generation of a genomic tiling array of the human major histocompatibility complex (MHC) and its application for DNA methylation analysis." *BMC Med Genomics* (2008). 1: 19.





- Tran, R. K., J. G. Henikoff, D. Zilberman, R. F. Ditt, S. E. Jacobsen and S. Henikoff "DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes." *Curr Biol* (2005). 15(2): 154-159.
- Trapnell, C., D. G. Hendrickson, M. Sauvageau, L. Goff, J. L. Rinn and L. Pachter "Differential analysis of gene regulation at transcript resolution with RNA-seq." *Nat Biotechnol* (2013). 31(1): 46-53.
- Tsukahara, S., A. Kobayashi, A. Kawabe, O. Mathieu, A. Miura and T. Kakutani "Bursts of retrotransposition reproduced in Arabidopsis." *Nature* (2009). 461(7262): 423-426.
- Vaughn, M. W., M. Tanurdzic, Z. Lippman, H. Jiang, R. Carrasquillo, P. D. Rabinowicz, N. Dedhia, W. R. McCombie, N. Agier, A. Bulski, V. Colot, R. W. Doerge and R. A. Martienssen "Epigenetic natural variation in Arabidopsis thaliana." *PLoS biology* (2007a). 5(7): e174.
- Vaughn, M. W., M. Tanurdzic, Z. Lippman, H. Jiang, R. Carrasquillo, P. D. Rabinowicz, N. Dedhia, W. R. McCombie, N. Agier, A. Bulski, V. Colot, R. W. Doerge and R. A. Martienssen "Epigenetic natural variation in Arabidopsis thaliana." *PLoS Biol* (2007b). 5(7): e174.
- Vielle-Calzada, J. P., J. Thomas, C. Spillane, A. Coluccio, M. A. Hoepfner and U. Grossniklaus "Maintenance of genomic imprinting at the Arabidopsis medea locus requires zygotic DDM1 activity." *Genes & development* (1999). 13(22): 2971-2982.
- Vining, K. J., K. R. Pomraning, L. J. Wilhelm, H. D. Priest, M. Pellegrini, T. C. Mockler, M. Freitag and S. H. Strauss "Dynamic DNA cytosine methylation in the Populus trichocarpa genome: tissue-level variation and relationship to gene expression." *BMC Genomics* (2012). 13: 27.
- Vongs, A., T. Kakutani, R. A. Martienssen and E. J. Richards "Arabidopsis thaliana DNA methylation mutants." *Science* (1993). 260(5116): 1926-1928.
- Vu, T. M., M. Nakamura, J. P. Calarco, D. Susaki, P. Q. Lim, T. Kinoshita, T. Higashiyama, R. A. Martienssen and F. Berger "RNA-directed DNA methylation regulates parental genomic imprinting at several loci in Arabidopsis." *Development* (2013). 140(14): 2953-2960.
- Wang, J., L. Tian, H. S. Lee and Z. J. Chen "Nonadditive regulation of FRI and FLC loci mediates flowering-time variation in Arabidopsis allopolyploids." *Genetics* (2006a). 173(2): 965-974.
- Wang, J., L. Tian, H. S. Lee, N. E. Wei, H. Jiang, B. Watson, A. Madlung, T. C. Osborn, R. W. Doerge, L. Comai and Z. J. Chen "Genomewide nonadditive gene regulation in Arabidopsis allotetraploids." *Genetics* (2006b). 172(1): 507-517.
- Wang, J., L. Tian, A. Madlung, H. S. Lee, M. Chen, J. J. Lee, B. Watson, T. Kagochi, L. Comai and Z. J. Chen "Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids." *Genetics* (2004). 167(4): 1961-1973.
- Wang, X., C. G. Duan, K. Tang, B. Wang, H. Zhang, M. Lei, K. Lu, S. K. Mangrauthia, P. Wang, G. Zhu, Y. Zhao and J. K. Zhu "RNA-binding protein regulates plant DNA methylation by controlling mRNA processing at the intronic heterochromatin-containing gene IBM1." *Proceedings of the National Academy of Sciences of the United States of America* (2013). 110(38): 15467-15472.



- Weber, M., J. J. Davies, D. Wittig, E. J. Oakeley, M. Haase, W. L. Lam and D. Schubeler "Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells." *Nat Genet* (2005). 37(8): 853-862.
- Weber, M., I. Hellmann, M. B. Stadler, L. Ramos, S. Paabo, M. Rebhan and D. Schubeler "Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome." *Nat Genet* (2007). 39(4): 457-466.
- Wolff, P., I. Weinhofer, J. Seguin, P. Roszak, C. Beisel, M. T. Donoghue, C. Spillane, M. Nordborg, M. Rehmsmeier and C. Kohler "High-resolution analysis of parent-of-origin allelic expression in the Arabidopsis Endosperm." *PLoS genetics* (2011). 7(6): e1002126.
- Wu, K., K. Malik, L. Tian, D. Brown and B. Miki "Functional analysis of a RPD3 histone deacetylase homologue in Arabidopsis thaliana." *Plant molecular biology* (2000). 44(2): 167-176.
- Xu, L., Z. Zhao, A. Dong, L. Soubigou-Taconnat, J. P. Renou, A. Steinmetz and W. H. Shen "Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in Arabidopsis thaliana." *Molecular and cellular biology* (2008). 28(4): 1348-1360.
- Yamamuro, C., D. Miki, Z. Zheng, J. Ma, J. Wang, Z. Yang, J. Dong and J. K. Zhu "Overproduction of stomatal lineage cells in Arabidopsis mutants defective in active DNA demethylation." *Nature communications* (2014). 5: 4062.
- Yang, C., F. Bratzel, N. Hohmann, M. Koch, F. Turck and M. Calonje "VAL- and AtBMI1-mediated H2Aub initiate the switch from embryonic to postgerminative growth in Arabidopsis." *Current biology : CB* (2013). 23(14): 1324-1329.
- Yi, H., N. Sardesai, T. Fujinuma, C. W. Chan, Veena and S. B. Gelvin "Constitutive expression exposes functional redundancy between the Arabidopsis histone H2A gene HTA1 and other H2A gene family members." *The Plant cell* (2006). 18(7): 1575-1589.
- Yu, A., G. Lepere, F. Jay, J. Wang, L. Bapaume, Y. Wang, A. L. Abraham, J. Penterman, R. L. Fischer, O. Voinnet and L. Navarro "Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense." *Proceedings of the National Academy of Sciences of the United States of America* (2013). 110(6): 2389-2394.
- Yu, X., L. Li, M. Guo, J. Chory and Y. Yin "Modulation of brassinosteroid-regulated gene expression by Jumonji domain-containing proteins ELF6 and REF6 in Arabidopsis." *Proceedings of the National Academy of Sciences of the United States of America* (2008). 105(21): 7618-7623.
- Zemach, A., M. Y. Kim, P. H. Hsieh, D. Coleman-Derr, L. Eshed-Williams, K. Thao, S. L. Harmer and D. Zilberman "The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin." *Cell* (2013). 153(1): 193-205.
- Zemach, A., I. E. McDaniel, P. Silva and D. Zilberman "Genome-wide evolutionary analysis of eukaryotic DNA methylation." *Science* (2010). 328(5980): 916-919.
- Zhang, H., Z. Y. Ma, L. Zeng, K. Tanaka, C. J. Zhang, J. Ma, G. Bai, P. Wang, S. W. Zhang, Z. W. Liu, T. Cai, K. Tang, R. Liu, X. Shi, X. J. He and J. K. Zhu "DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV." *Proc Natl Acad Sci U S A* (2013). 110(20): 8290-8295.



- Zhang, X., Y. V. Bernatavichute, S. Cokus, M. Pellegrini and S. E. Jacobsen "Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*." *Genome biology* (2009). 10(6): R62.
- Zhang, X., J. Yazaki, A. Sundaresan, S. Cokus, S. W. Chan, H. Chen, I. R. Henderson, P. Shinn, M. Pellegrini, S. E. Jacobsen and J. R. Ecker "Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*." *Cell* (2006). 126(6): 1189-1201.
- Zhao, Y., S. Xie, X. Li, C. Wang, Z. Chen, J. Lai and Z. Gong "REPRESSOR OF SILENCING5 Encodes a Member of the Small Heat Shock Protein Family and Is Required for DNA Demethylation in *Arabidopsis*." *The Plant cell* (2014).
- Zheng, X., O. Pontes, J. Zhu, D. Miki, F. Zhang, W. X. Li, K. Iida, A. Kapoor, C. S. Pikaard and J. K. Zhu "ROS3 is an RNA-binding protein required for DNA demethylation in *Arabidopsis*." *Nature* (2008). 455(7217): 1259-1262.
- Zhong, X., C. J. Hale, M. Nguyen, I. Ausin, M. Groth, J. Hetzel, A. A. Vashisht, I. R. Henderson, J. A. Wohlschlegel and S. E. Jacobsen "Domains rearranged methyltransferase3 controls DNA methylation and regulates RNA polymerase V transcript abundance in *Arabidopsis*." *Proc Natl Acad Sci U S A* (2015). 112(3): 911-916.
- Zhu, J., A. Kapoor, V. V. Sridhar, F. Agius and J. K. Zhu "The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in *Arabidopsis*." *Current biology : CB* (2007). 17(1): 54-59.
- Zilberman, D., D. Coleman-Derr, T. Ballinger and S. Henikoff "Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks." *Nature* (2008). 456(7218): 125-129.
- Zilberman, D., M. Gehring, R. K. Tran, T. Ballinger and S. Henikoff "Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription." *Nat Genet* (2007). 39(1): 61-69.
- Zubko, E., M. Gentry, A. Kunova and P. Meyer "De novo DNA methylation activity of methyltransferase 1 (MET1) partially restores body methylation in *Arabidopsis thaliana*." *The Plant journal : for cell and molecular biology* (2012). 71(6): 1029-1037.